

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

<http://go.warwick.ac.uk/wrap/66605>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

# **Alienation and the Sciences of Mind**

## **Understanding Schizophrenia without Cognitivist Theory**

**Richard Gipps**

**Dissertation Submitted for the degree of**

**Doctor of Philosophy**

**Warwick University  
Department of Philosophy**

**September 2001**

## **Contents**

<b>Summary</b>	<b>p. 5</b>
----------------	-------------

### **Part 1. A Sense of the Issues**

#### **Ch. 1. Diagnosing the Disorders.**

##### **1. Introduction: Cognitivism, Schizophrenia, Alienation.**

i. Why Cognitivism and Schizophrenia?	p. 7
ii. Grammar, Epistemology, and Psychological Theory.	p. 9
iii. Alienation and Schizophrenia.	p. 12

##### **2. 'Newton's Error'.**

i. Newton on Absolute Motion.	p. 16
ii. The Metaphysical Roots of 'Newton's Error'.	p. 19

##### **3. The Agential Background.**

i. Grammar, Necessity, and the Background.	p. 23
ii. The Structure of the Background.	p. 25

<b>4. On What is to Follow.</b>	<b>p. 32</b>
---------------------------------	--------------

### **Part 2. Understanding Mindedness**

#### **Ch. 2. Our Inner Lives.**

##### **1. Introduction: On the Mind as 'Inner'.**

i. Preliminary	p. 35
ii. What is Cognitivism?	p. 36

##### **2. Perception as Input.**

i. Helmholtzian Problematics in the Psychology of Perception.	p. 40
ii. Fodor vs. Ryle on Perceptual Processes.	p. 44
iii. Fodor vs. Ryle on 'Perception Recipes'.	p. 50

##### **3. Thought as Representation.**

i. Mental Models and Psychological Explanation.	p. 55
ii. Tacit Knowledge.	p. 61
iii. Fundamental Diagnosis and Cognitivist Theories of Meaning.	p. 66
iv. Summing Up.	p. 73

### **Ch. 3. On Reading Others and Expressing Ourselves.**

- 1. Introduction.** p. 74
- 2. On Reading Others.**
  - i. Wittgenstein on Other Minds. p. 75
  - ii. The Case of Dreaming (Fodor vs. Wittgenstein). p. 80
  - iii. Fodor on *Inference to the Best Explanation*. p. 88
    - a. *Explanation* in Folk Psychology. p. 89
    - b. *Inference* in Folk Psychology. p. 93
  - iv. The Defeasibility of Behavioural Criteria. p. 99
  - v. The Psychological Character of Behaviour. p. 100
  - vi. The Desirability of Explanation. p. 103
  - vii. Summing Up. p. 106
- 3. On Expressing Ourselves.** p. 108
  - i. Problems with Cognitivism. p. 110
  - ii. Avowalism. p. 115
  - iii. Summing Up. p. 120

## **Part 3. Understanding Psychosis**

### **Ch. 4. Cognitive Theories of Schizophrenic Fragmentation.**

- 1. Introduction: Mind and Mechanism in Schizophrenia and Cognitivism.**
  - i. The Influencing Machine in Schizophrenia. p. 122
  - ii. The Influencing Machine in Psychology. p. 124
  - iii. The Structure of the Argument. p. 127
- 2. Schizophrenic Motivation and Action.**
  - i. Frith's Cognitive Theory of Schizophrenia. p. 131
  - ii. Failure of Action: The Theory. p. 131
  - iii. The Explanatory Power of Frith's Action Model. p. 134
  - iv. The Predictive Scope of Frith's Action Theory. p. 138
- 3. Schizophrenic Affect and Expression.**
  - i. Frith's Cognitive Theory of Emotion. p. 142
  - ii. Critique of the Cognitive Theory of Emotion. p. 143
- 4. Schizophrenic Thought and Language.**
  - i. Incoherence of Speech. p. 145
  - ii. Thought Disorder. p. 147
  - iii. Conclusion. p. 152



## **Ch. 5. Cognitive Theories of Schizophrenic Self-Estrangement.**

- 1. Introduction.** p. 154
- 2. Philosophical Theories of Psychotic Experience.**
  - i. Immunity to Error? p. 155
  - ii. The Psychology of 'Introspective Alienation' and the Rhetoric of 'Self-Consciousness'. p. 159
  - iii. Thought and Agency: The Action Analogy. p. 164
  - iv. Thought and Ownership: The Attribution of Thoughts. p. 170
- 3. Cognitive Psychological Theories of Psychotic Experience.**
  - i. Psychological Theories of Hallucination. p. 175
  - ii. Critique of 'Output' Theories of Hallucination. p. 177
  - iii. Cognitive Theories of Thought Insertion. p. 180
  - iv. The Alienated Roots of Cognitive Theories of Introspective Alienation. p. 183

## **Part 4. The Psychotic Core**

## **Ch. 6. Cognitive Theories of Schizophrenic World-Estrangement.**

- 1. Introduction: The Phenomenology and Epistemology of Delusion.** p. 192
- 2. Preliminary Remarks on the Character of Delusion.**
  - i. Internal and External Characteristics. p. 196
  - ii. Primary and Secondary Delusions, or 'True' Delusions and 'Delusion-Like Ideas'. p. 202
- 3. Cognitive Theories of Delusion as Failures in Rationality.**
  - i. Empirical Theories and Evidence. p. 207
  - ii. Critique of Suppressed Premises. p. 210
- 4. Delusion as Rational Response.**
  - i. Maher's Theory of Delusions as Normal Theories. p. 214
  - ii. Critique of Maher's Theory. p. 216
- 5. Conclusion** p. 221

## **Ch. 7. The Grammar of the Divided Self.**

- 1. Introduction: Schizophrenia and the Fragmentation of the Background.** p. 224
- 2. Ontological Fragmentation in Schizophrenia.**

- i. John Hyman's Treatment of Blindsight as a Template for Understanding Schizophrenia. p. 229
- ii. Applying the Template to Schizophrenia. p. 231

### **3. Schizophrenic Self-Estrangement.**

- i. Wittgenstein's 'Secondary Sense' as a Template for Understanding Thought Insertion. p. 238
- ii. Applying the Template. p. 241

### **4. Delusion: Epistemological Alienation in Schizophrenia.**

- i. Jaspers and *das Wahnproblem*. p. 245
- ii. The Intrusive Knowledge of Meaning. p. 246
- iii. Change in the Personality 1: Spitzer's Formulation. p. 249
- iv. Change in the Personality 2: Fulford, Campbell and Eilan. p. 252
- v. On the Road to the Psychotic Core: The Praxical Foundations of Delusion. p. 255
- vi. Subpersonal Dynamics and the Rehabilitation of Cognitive Neuropsychology. p. 259

## Summary

This dissertation combines a philosophical critique of cognitive theories of schizophrenia with an alternative theorisation of the alienated states of mind met with in this illness.

Whilst schizophrenia constitutes the *embodiment* of an estranged subjectivity – a profound alienation from the world, other people and oneself, cognitivism provides us (so the argument goes) with a fundamentally alienated *conception* of mind. Because it tacitly occupies such a perspective, it can easily appear as if cognitivism has provided a suitable framework for theorising schizophrenia. But insofar as it fails to adequately depict our engagement with others and with the world, or explain the integrity of our personality, cognitivist theory fails to provide a framework within which a psychotic breakdown in such engagement or integrity can be understood.

Because cognitivism provides a conception of mind as a self-contained inner realm disengaged from the social world and natural environment, it gives rise to a series of metaphysical and epistemological problematics. These inspire theories and explanations of causal mechanisms and epistemic faculties which serve to reconnect the subject with that (the world, other subjects, their own minds) from which they have theoretically been alienated. Without such a disengaged perspective, however, the apparent need for theory and explanation does not arise. What is rather required is a more adequate conception of mind as intrinsically relational, and of individual mental contents as constitutively embedded within a Background both of other such contents and of meaningful bodily behaviour.

Cognitive accounts of schizophrenia explain the illness in terms of breakdowns within the aforementioned causal mechanisms and epistemic faculties. These mechanisms and faculties however are mythical and are in any case posited to perform an unnecessary task. A better understanding of schizophrenia can be had in terms of fractures within the aforementioned Background structures of meaning.

# **Part 1**

## **A Sense of the Issues**

## Ch. 1. Diagnosing the Disorders

### 1. Introduction: Cognitivism, Schizophrenia, Alienation – A Sense of the Issues

#### i. Why Cognitivism and Schizophrenia?

The seven chapters to follow combine a critique of cognitivism as a philosophical theory of mind with an inquiry into psychological theories of schizophrenia. An obvious question to ask is why I have treated such topics together, a question which allows for a preliminary elaboration of the issues.

Gaining a psychological understanding of schizophrenia can be argued to be more than just one amongst many of psychology's goals. Most of us feel fairly comfortable in our understanding of rational thought and action, and with the mild inflections of and distortions to these introduced by moods, temperament and personality. Psychosis in general and schizophrenia in particular, however, offer a far greater challenge. Our normal 'folk-psychological' techniques for understanding people just don't seem to be able to get much of a purchase on the utterances and actions of those suffering from severe mental illness. Because the minds and lives of such people can appear impenetrable and frightening it is natural to see if our chance of understanding them can be improved by supplementing our intuitive folk-psychological know-how with the theoretical modes of knowledge provided by psychology. In so far, then, as psychological science exists to help us understand what we do not already understand about the mind, and in so far as schizophrenia represents one of the most intuitively unintelligible mental conditions, developing an understanding of schizophrenia represents psychology's greatest challenge.

Today psychology is largely approached as a 'cognitive science'; behaviourism's fascination with 'stimulus' and 'response' has been replaced by a form of inquiry interested in the 'cognitive states' and 'cognitive processes' which, we are told, intervene between the two<sup>1</sup>. This interest in these 'intervening variables' has been widely portrayed as a re-introduction of the actual subject matter of psychology – i.e. the *psyche* or mind – back into the discipline, a subject matter which behaviourism, labouring under dubious positivistic influences, had misguidedly attempted to disappear. The study of psychological

---

<sup>1</sup> C.f. Jerry Fodor's *Psychological Explanation* [henceforth *PE*], ch.2.

disorders has likewise taken a 'cognitive turn, and the principle accounts of psychosis – especially of schizophrenia – on the theoretical market today employ cognitive models of the mind<sup>2</sup>.

To try and understand schizophrenia, then, I turned to cognitive psychology and investigated the models on offer. To cut a long story short: I was disappointed. The models of schizophrenia on offer did not seem to capture the nature of psychotic experience or delusional thought. And the difficulties did not appear to be in the details either: what seemed off key was the *form* (and not the content) of the explanations on offer. They just didn't seem to be the right *kind* of explanations for what it might mean to lose contact with reality or to have thoughts and experiences which are nearly unthinkable or unimaginable. In fact, to the extent that the psychotic symptoms were explained, it seemed that it was only by first tacitly assimilating them to more normal and manageable mental phenomena; without this assimilation the cognitive models just couldn't get off the ground.

Even this tacit assimilation did not however seem to get to the heart of the problem. The root of the failure to get to grips with the failed reality contact and disruption of mindedness manifest in schizophrenia appeared to stem from a failure of the psychological models provided by cognitive science to adequately theorise reality-contact and the conditions of mindedness themselves. It is for this reason, then, that this dissertation concerns itself with both the theoretical foundations of cognitive psychology and with schizophrenia. The task is to rethink these foundations and to see whether or not such a rethink of the nature of our foundation of our contact with the world and of the integrity of our mindedness can provide for a more adequate understanding of schizophrenia.

It has to be admitted however that, in the space provided, what follows is concerned more with critique and less with the explanation of schizophrenia. (In fact it is only in the final chapter that any positive kind of characterisation of psychotic symptomatology is developed.) One sort of excuse for this can be found in the *philosophical* rather than *psychological* remit of the dissertation. What is important here is getting clear about the questions, examining the presuppositions, and laying the foundations for a psychology of schizophrenia – and not the empirical details. Another reason for the imbalance stems from a growing conviction that it may not be from *explanation* that the *psychological understanding*

---

<sup>2</sup> Such as Christopher Frith's *The Cognitive Neuropsychology of Schizophrenia* [CN].

sought by someone looking to comprehend what it means to suffer from schizophrenia is to be provided. I shall now explain what I mean by this.

## ii. Grammar, Epistemology, and Psychological Theory

As already mentioned, cognitive theory aimed to replace behaviourist approaches to psychology by introducing mind as an intervening variable between behaviourism's stimulus and response, theorising the mind in terms of inner processes mediating between perceptual 'input' and behavioural 'output'. In so doing it made itself vulnerable to those mid-twentieth-century philosophical objections directed toward *mentalist* conceptions of the mind, that is, made itself vulnerable to those arguments against the conception of mind as an *inner realm populated by inner states and inner processes* developed by philosophers such as Ludwig Wittgenstein and Gilbert Ryle.<sup>3</sup> It was not however merely in the content of the explanations of mind that cognitive theory made itself vulnerable, but also in the very fact that explanations were given, for a striking feature of such mid-century philosophy was the prohibition it proclaimed on theorising and explanation itself.

Cognitive theory has, historically, not shown itself to be too worried about such objections. On the one hand it tended to see Ryleian or Wittgensteinian objections to mentalism as simply another manifestation of the positivism and behaviourism that they aimed to overthrow. On the other hand the objection to theorising and explanation was seen as an anti-scientific prejudice and an unprincipled whim which could only hold back the development of all-important scientific theories of mind<sup>4</sup>. This I think was a mistake. Whilst the argument must wait until Part 2 the rest of this section will provide a schematic outline; the following sections will introduce some of the philosophical apparatus employed and psychological themes pursued throughout the dissertation.

Wittgenstein and Ryle's philosophy was first and foremost *linguistic* philosophy – it concerned itself with the 'logical', 'grammatical' or 'categorical' structure and character of our various 'discourses' or

---

<sup>3</sup> Wittgenstein, *Philosophical Investigations* [PI]; Ryle, *The Concept of Mind* [CM]. For an application of their arguments to the programme of cognitive science see Button et al.'s *Computers, Minds and Conduct*.

<sup>4</sup> C.f. Fodor's *PE*.

'language-games', and it was from the perspective of the understanding generated by such a perspective that traditional problems in epistemology and metaphysics were tackled<sup>5</sup>. A key concern was to look at the *different* ways in which the words making up our ordinary (pre-philosophical) language gained the meanings they had; the argument was that theorists of the past had failed to notice the radical logical heterogeneity of the language-games which constituted our natural languages, assimilating diverse categorical structures to a single familiar pattern. As I see it the most typical of such assimilations involved construing: i) the grammar of *all utterances* on the model of *assertions*, ii) the *grammatically defined objects* of such utterances on the model of the grammar of medium-sized *physical objects* - and as *occupying some kind of space* or 'realm', and iii) the *validity* of all such utterances in the mouth of their pronouncer as resting on an *epistemic or quasi-perceptual ground*.

When this insight is applied to the mind, the typical objection is that epistemologists and psychologists tend to construe the mind as some kind of inner realm or space or container and tend to construe the 'contents' of the mind (desires, hopes, fears, intentions, beliefs etc.) as, literally, inner contents of this inner realm. Correspondingly, utterances detailing our own thoughts and feelings are viewed as typically assertions, assertions expressing beliefs which we are able to arrive at by the use of a kind of quasi-perceptual faculty of 'inner sense'. Utterances describing the mental contents of others are thought of as assertions about an inner realm, assertions expressing beliefs arrived at from inferences from outer behaviour. Furthermore, Ryle's epithet for the 'category mistake' made by the mentalist - embodied in their doctrine of the 'ghost in the machine' - reveals that their mistake was thought to lie not merely in the construal of mind as an inner (ghostly) realm, but also in the construal of behaviour and action as something intrinsically non-mental, mechanistic and 'outer' - as mere movement.<sup>6</sup>

Seen in this light the cognitivist's desire to supplement the behaviourist's talk of stimulus and response, or as they themselves would put it, input and output, with something genuinely psychological is all well and good. The trouble however is that, far from undoing the category mistake at the root of the behaviourist program, it is simply recapitulated by the cognitivist, for the mind with which they re-inject life into the behaviourist's body is nothing other than the mentalist's inner realm of inner states and inner

---

<sup>5</sup> C.f. Freidrich Waismann's *Principles of Linguistic Philosophy*, part 1.

<sup>6</sup> Some of these general objections are collected by Sprague in his *Persons and their Minds*.



processes. Far from Wittgenstein and Ryle's critique of mentalism being based in a behaviourist mind-denying prejudice and in a narrowed view of the available options, it is the mentalist – and by implication the cognitivist – who narrows the field and denies the distinctiveness of mind, assimilating the grammar of both mind and behaviour to that of physical bodies and physical spaces.

What then of the above-mentioned anti-theoretical injunctions frequently associated with Wittgenstein's philosophy? Some of this may be a function of personality, and of course even a purely descriptive overview of the categorical structure of various discourses could be – albeit at the risk of pretension – called a theory<sup>7</sup>. Nevertheless there is also reason to think that Wittgenstein's methodological concerns with proper philosophical procedure are not disconnected from the metaphysical issues at stake. The argument is that a range of typical philosophical 'how?' questions, explanations and theories are only asked and produced in response to a perplexity that itself arises once a host of assumptions have already been tacitly made about the nature of the categorical structure of the discourse in question.

This can be seen once the effects of the mentalist's characterisation of the psychological domain of discourse are calculated. If the mind is theorised as an inner realm, constitutively cut off from the outer world and even from bodily behaviour, there will be a need for epistemological theories which make it clear how epistemic contact with the world can be achieved. If our avowals of our own mental contents are thought of as assertions about what is going on within us there will be a need for a theory of a faculty of inner sense to show how we come by our beliefs about what is to be found there. If behaviour is a merely outer phenomenon constitutively divorced from the mind there will be a need for theories explaining how we are able to come by our judgements concerning the minds of others. If meaning is to be accounted for in terms of inner mental representations, then there will be a need for a theory explaining how these inner representations can stand for objects in the world.

If, then, mind and meaning are theorised in terms of a self-contained and alienated inner realm, the epistemologist will have their job cut out developing theories both in order to fend off scepticism (scepticism that we can ever really know what is going on in the 'outer' world, or in the minds of others) and to sketch the general character of our engagement with the world, with others, and even with the

---

<sup>7</sup> As in the 'descriptive metaphysics' found in Strawson's *Individuals*.

contents of our own minds. Once the outline is developed the job can be handed over to the empirical psychologist in order to fill in the details of precisely what mechanisms are used to implement the epistemic contact. Change the picture, however, view mind and meaning as not hidden behind supposedly 'outer' bodily behaviour but as fully imminent within action which is itself intrinsically intentional, and the need for the epistemological theories and their psychological elaborations simply drops away. If this is right then Wittgensteinian prohibition on theory can be understood, not as an *a priori* stricture on legitimate philosophising, but as an expression of the view that epistemological theorising tends to be predicated upon certain assumptions about the categorical structure of the language-games in question. Undo these assumptions, provide a descriptive account of the grammar of the discourse domain, and the 'how' questions, questions which seemed to demand a theoretical answer in terms of epistemic faculties and cognitive processes, no longer arise.

Returning to the question of schizophrenia: the argument of Part 3 of this dissertation will be that cognitive theories tend to explain the disorder in terms of breakdowns in epistemic faculties or causal mechanisms, mechanisms and faculties which, on the argument presented, are simply imaginary devices performing imaginary functions. It is for this reason that the cognitive accounts seem to fail to adequately theorise the disorder. Part 4 will also argue that the purely descriptive account of mind developed in Parts 1 & 2 contains the resources for developing an understanding of the psychopathology without the need for explanation.

### **iii. Alienation and Schizophrenia**

Another central theme of the dissertation is that of *alienation*. As concerns cognitivism the effects of alienation have already been suggested at: in seeing the mind as an *inner* realm the subject is portrayed as constitutively cut off from their environment and from others. Furthermore in supposing that epistemic access is required to one's own mind own the subject retreats not merely behind the body but also behind

their own mind, becoming now an onlooker to impersonally characterised *events, processes* and *states* occurring there.<sup>8</sup>

Alienation characterises not merely the cognitive theories but also the content matter of such theories: schizophrenia was itself once known as a condition of *profound alienation*; psychiatrists too used to be known as alienists. A person suffering from schizophrenia will often feel cut off from their own body, or feel that their actions, feelings and thoughts are not their own, or experience as foreign voices which can be shown to correspond to their own whispered utterances. The world too may feel like a foreign country to someone with schizophrenia, feel like a glassy unreal realm, and other people can come to appear as if automata, mindless or dead. The subject with schizophrenia is often said to have lost touch with reality and to be trapped and drifting within their own private autistic world of dream and phantasy.

The parallels between the alienated *conception* of the subject provided by cognitivism and the alienated form of subjectivity *embodied* by schizophrenia are certainly striking, and have been explored to marvellous depths by the clinical psychologist Louis Sass.<sup>9</sup> This however is not my concern, which is rather to show not only why cognitive theories of mind and of madness ultimately fail, but also why they appear to *succeed*. I shall now explain what I mean by this.

Consider first the apparent success. Cognitivism provides us with an account of mind as not constitutively but rather merely causally bound up with its environment and with other minds, and as internally constituted by discrete modules that again are in merely causal interaction. And schizophrenia is a condition in which the mind – *in some sense* – can be said to be both internally fragmented and also

---

<sup>8</sup> Mentalism not only tends to think of mind as an inner realm, but objectifies it, depriving it of its constitutive subjectivity. Not only are psychological phenomena now predicated of *the mind* rather than the body, but a rhetoric of impersonal forms of perceptual and cognitive verbs develops: 'perceivings', 'believings', 'notings', 'rememberings'. (For examples of bizarre gerunds formed from psychological verbs see Donald Davidson's *Mental Events*.) The alienated subject retreats behind the mind and looks in at these objective goings on.

<sup>9</sup> *Madness and Modernism*; see also *The Paradoxes of Delusion*. Mechanistic mentalism (that which I call 'cognitivism') can be considered to be modernism's theory of mind. Sass not only considers modernism's objectification of the mind but also its subjectivising of the world, and uses both to develop a clinical understanding of the delusional world of schizophrenia.

cut off from the world. What, then, could be more natural than to locate the internal fragmentation and external alienation of the schizophrenic mind in the causal links postulated by the cognitive theories?

There has nevertheless been, as any reader of the history of psychopathology will recognise, hardly any development in our psychological understanding of the nature of schizophrenia since Bleuler's (1911) monumental *Dementia Praecox: Or the Group of Schizophrenias*, or Jaspers' (1913) equally impressive *General Psychopathology*. In fact in many ways our current textbooks fail to capture both the wealth of phenomenological detail presented by these earlier texts and the subtlety of their theoretical formulations. And this, I would argue, is in large part the consequence of the baleful influence of mechanistic mentalism, and the disengaged conception of mind that it brings with it. For whatever the *apparent* plausibility of explaining the fragmentation and dislocation of the schizophrenic mind in terms of cognitive disconnections - between self and mental content, thought and language, or, for example, between intention and action, - the fact remains that our ability to understand what it is to be psychotic has largely remained static, and that the central phenomenology of psychosis - for example, delusions and delusional perception - and the commonly recognised conceptual problems faced by psychiatry - as to the nature of delusion, for example, or of insight, the grounds for psychiatric diagnoses, and the validity of concepts like 'schizophrenia' - remain recalcitrant to cognitivist analysis.

The reasons for the apparent success and actual failure are complicated, and will by and large have to emerge from the argument of the following chapters. The structure of the argument can however be advertised in advance. In the process we can begin to see how it can be maintained - as I did above - both that that cognitive explanations of schizophrenic symptoms only 'work' by assimilating such phenomena to conditions of normality, and also that the cognitive theorisations of normality are themselves inadequate.

A consequence of cognitivism's objectivised conception of mind, as has been noted, is not only a certain view of mental contents but also an alienated conception of the mind and of the human subject. Cognitivist explanations of mind are supposed to explain the functions of our subjectivity and our agency, but on the argument to be developed *they actually presuppose them*. They presuppose them because the phenomena in terms of which the mind is theorised by the cognitivist require a further subject to interpret them, or to act upon them, or to perceive them. The cognitivist might for example

attempt to explain perception in terms of the having of inner images, but such images will require an inner subject to perceive them. Bodily actions might be explained in terms of acts of will, but here too the explanation presupposes that which is to be explained. What a subject *means* by their 'outer' utterances or writings may be theorised in terms of inner representations, but representations require a subject to interpret them or use them in a certain way before they can be said to mean anything: once again we are back where we started. As already noted a common cognitivist tendency is to explain avowal in terms of assertion, but an assertion is itself an avowal of a belief. Again and again the cognitivist presupposes what they are supposed to be explaining in their explanations. It is for this reason that it is misleading to think that cognitivism provides us with a merely reductive account of mind; all the time a tacit, implicit and illicit subject retreats into the background in order to be the inner observer and interpreter of representations and initiator of inner actions<sup>10</sup>.

Furthermore, this subject that retreats into the background is *a fully sane subject*, and it is, when considering a subject suffering from schizophrenia, by tacitly putting ourselves into the shoes of the alienated inner subject that we gain the impression that we are able to understand the symptomatology. It is for this reason that one all too often gets the impression, when reading cognitive theories of psychotic illness, that *what are being discussed are the perfectly rational responses of perfectly sane subjects*. It is also for this reason that the cognitive theories fail: they give an explanation where none is required, and the explanation that they do give itself presupposes what it is trying to explain. Cognitivism aims to explain: our practical engaged activity, our perception and understanding of the world, our fundamental rationality, our linguistic competence, our capacity to read the mental lives of others, all in terms of a set of non-praxical and fairly intellectual contents and competencies: unconscious inferences, tacit knowledge, representational knowledge, mental representations, introspection, theories of mind etc. But such disengaged cognitive operations, so the argument goes, actually *presuppose, and so can't explain, the very praxical capacities for which a theoretical account is supposed to be given*.<sup>11</sup>

---

<sup>10</sup> Daniel Dennett provides an approximately similar scheme in his book *Consciousness Explained*, which is primarily designed to dispel the illusion that, after the brain has performed all of its information processing on incoming sensory material, it must send the processed information to a central collection point for it to be presented to consciousness. Dennett refers to this as the myth of the Cartesian Theatre.

<sup>11</sup> Cf. Julia Tanney, *Playing the Rule-Following Game*.

A final theme that will be of importance in this dissertation is that of the 'background'. The argument will be that in theorising some or other discourse domain along the grammatical lines of that concerned with physical bodies, a whole host of complex and ramifying conditions of intelligibility for the concepts of the discourse domain in question are simply overlooked. In the case of the background implicated in psychological discourse – which I shall refer to (with a capital 'B') as the *Background* – a whole range of complex situational and frequently defeasible preconditions for the ascription of propositional attitudes, thoughts, moods, actions and perception are typically ignored. On the one hand this is not surprising for the Background, whilst pervasive, is hard to bring into focus as foreground. On the other hand it is disastrous, for not only does ignoring it lead to a failed theorisation of mind and the development of spurious epistemological problematics, but it also makes it impossible to understand schizophrenia, a disorder which, if the thesis of Part 4 is correct, must itself be understood in terms of splits and reorganisations within the Background.

In the following section I shall explicate the general notion of background, and of how its being overlooked results in the development of unnecessary theories which in any case presuppose what they aim to explain, by means of an uncontroversial analogy from physics. In section 3 I shall describe the agential background – the Background – in some more detail.

## **2. 'Newton's Error'**

### **i. Newton on Absolute Motion**

To label the philosophical confusion to be considered as 'Newton's error' is perhaps somewhat mischievous, in as much as Newton, maybe the greatest of all scientists, is best remembered for his outstanding contributions and not for his mistakes. But the preface to Book 1 of the *Philosophiae Naturalis Principia Mathematica* contains so precisely the form of confusion on which I wish to focus – both in as much as it overlooks the central rôle of what I am calling the 'background', and in as much as this leads to the proliferation of metaphysical and epistemological theory where none in fact is required

or possible - contains so precisely the relevant confusion that the label is, even if mischievous, strikingly apt.

In the *Scholium* following the *Definitions* prefaced to the first book, Newton claimed that we must distinguish merely relative from 'true' or absolute motion. For our everyday purposes it may be adequate to calculate the motion of objects by noting their passage over the earth, but we should not, he suggests, forget that if we wish to determine their 'actual', 'true' or 'absolute' motion, the motion of the earth itself should be included in the equation.

As well as making the distinction between absolute and relative motion, Newton also distinguishes 'absolute, true, and mathematical' time from 'relative, apparent, and common time': the first 'from its own nature, flows equably without relation to anything external', whereas the latter 'is some sensible and external (whether accurate or unequable) measure of duration by the means of motion, which is commonly used instead of true time.' He similarly differentiates between 'absolute space' and 'relative space': the former 'in its own nature, without relation to anything external, remains always similar and immovable', whereas the latter 'is some movable dimension or measure of the absolute spaces; which our senses determine by its position to bodies'. Absolute velocity is accordingly to be ascertained by dividing a distance travelled in absolute space by the period taken, in absolute time, for the trajectory to be made.

Newton gives us an example of a sailor moving on a ship.

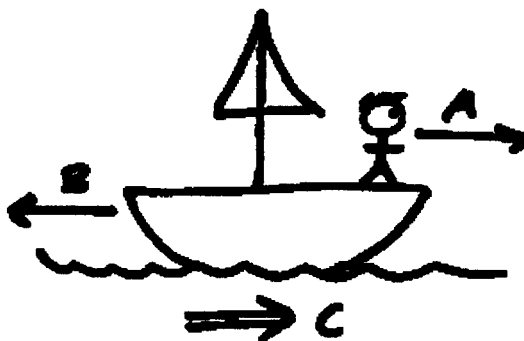


Fig 1.1

We might think that the velocity of the sailor (A) is to be had by simply measuring the rate of his progress along the ship (travelling east with a velocity of 1 unit, in Newton's example). But we shouldn't

forget to take into account the motion of the ship itself (B). Perhaps we measure the movement of the ship over the sea - the ship moves west over the sea with a velocity of 10 units. So the sailor moves west at a velocity of 9 units. But once again, we should consider that the earth itself (and therefore the sea) might be moving. Newton gives the velocity of the earth (C), as it travels eastwards through space (!), as 10,010 units. The sailor's true velocity, therefore, is  $10,010 - 10 + 1 = 10,001$  units to the east. Determining velocity C is no easy matter, but according to Newton this is merely an *epistemic* difficulty ('because the parts of that immovable space, in which these motions are performed, do by no means come under the observation of our senses'). The epistemic difficulty is not however insurmountable; by an examination of the 'causes, effects, and apparent differences' of the motion of bodies we can make out their 'true' motions; in fact, as he says, 'to this end it was that I composed' the *Mathematical Principles*.

Luckily, however, Newton's book is *not* really concerned with the possibility of that derivation, for the notions of absolute space, motion and time are run through with conceptual confusion. The difficulty in determining 'true' motions is not epistemic; rather: there are *no such movements to determine*. Although Sir Isaac is embarked on the venerable seventeenth and eighteenth century project of separating mere appearances from reality in the name of science, the separation of the *concept* of motion from that of the relative positions of bodies leads - not to empirical exactness but - to metaphysical nonsense.

Now Newton was surely right to point out that our everyday judgements of motion tacitly take the earth as a frame of reference. But from this observation he drew precisely the wrong conclusion. To be sure, our everyday talk of the motion of this or that object makes no explicit mention of a frame of reference; we go on, it could be said, as if motion were an intrinsic (non-relative) property of terrestrial objects. But from this observation - of the fact that our everyday discourse tacitly depends upon a background framework, but that we make no mention of this - Newton concluded that the face-value appearance of intrinsicity or absoluteness in the common discourse of motion was in fact its *true character*, and that the frame of reference was ultimately something that should be factored out, by including the ('absolute') motion of the frame within the calculation of motion. And in order to effect this he hypostasised the concept of space so that space becomes not the distance away from or between objects, or the dimensions of height, width and breadth, but rather a kind of 'container' of such objects.



Contrast with this procedure the moral Einstein drew from the same initial observation: 'The earth's crust plays such a dominant rôle in our daily life in judging the relative positions of bodies that it has led to an abstract conception of space which certainly cannot be defended. In order to free ourselves from this fatal error we shall speak only of 'bodies of reference', or 'space of reference'.<sup>12</sup> Whereas Newton assumed that what goes unnoticed and taken for granted in our everyday judgement must be exorcised, Einstein draws the opposite moral, that this tacit but essential background must be made visible precisely so that we will not fall into hopeless conceptions of space that attempt to do without it.

## ii. The Metaphysical Roots of 'Newton's Error'

Looking a bit more closely at Newton's rhetoric allows us to see the process by which spatio-temporal concepts are hypostasised and the (would-be) scientific program of determining absolute motions initiated. So, the earth is described as 'truly' moving through the 'parts' of 'immovable' space, the velocity of its 'passage' being a function of its motion over time, a time which 'from its own nature, flows equably without relation to anything external.' By treating space as a container, as being immovable, and as having parts ('Place is a part of space which a body takes up'), and by treating time as a flowing stream ('Absolute ... time, of itself, and from its own nature, flows equably without relation to anything external')<sup>13</sup>, Newton is able both to i) *deny that background preconditions are required for talk of motion whilst* ii) *at the same time covertly smuggle in the very same preconditions*. That this manoeuvre is illegitimate can be revealed by considering a) how the grammatical home of such concepts as *part* or *container* is one involving the description of *relations between objects*, b) how the notion of a *stream* or *flow* itself makes reference to the movement of bodies *over time* (and so can hardly be used in an account that is supposed to operate independently of such restrictions), c) how the notion of an *equable* flow is introduced without allowing anything (any natural oscillation) to even count as the measure of such equality, and d) how space and time cannot themselves be said to be moving *or*

<sup>12</sup> Albert Einstein, *The Meaning of Relativity* p.3

<sup>13</sup> Both cases conform to our earlier description of the characteristic form of philosophical dubosity described above (in section 1) – they characterise the grammar of space and time as akin to that of material objects.

'immovable'/'immutable'<sup>14</sup> In short, the condition of the possibility of the use of terms such as 'motion' and 'space', a condition labelled here as the 'background', is the condition of the impossibility of the explanatory projects envisaged by the (meta-)physicist.

The procedure can be understood in part in terms of the metaphilosophical considerations introduced in the above section. That is to say, the hypostatisation of the key concepts – the removal of 'space', 'time', 'motion' from their background intelligibility-conferring context – is accompanied by a misunderstanding of the categorical nature of spatio-temporal language games. And as is so often the case, the categorical structure is mistaken for that which structures empirical discourses about spatio-temporal *objects* and their motions. Precisely by these means, the preconditions for talk of motion (possessed by these ordinary discourses and manifest in our ordinary empirical claims) are smuggled back into the project, making it appear more plausible than it possibly could be. (Anthony Kenny has designated this general error the 'homunculus fallacy' when it is perpetrated in the psychological sciences.)

Let us now consider the consequences of Newton's overlooking of the *constitutive* rôle played by the background. Firstly, note that Newton's scheme is highly *Platonic*. That is to say, everyday motions, lengths and durations are treated as mere Appearances, ultimately to be accounted for, given a further explanation in terms of, a Platonic realm of Absolutes the existence of which is independent of the

---

<sup>14</sup> *Contra* paragraph 3 of *Scholium* section IV. It might be objected that the equable flow of absolute time, the absolute immovability of absolute space etc are to be understood with reference to *themselves only*; they act as *their own frame of reference*. This is however completely uninformative: in saying that something is equal to itself, one says nothing at all (except by way of ruling out a *logical* impossibility). It is noteworthy how frequently and how deeply the idea that something can be understood by reference to itself – and not in terms of contrasts with other terms of the conceptual scheme – runs in metaphysical texts. The private language argument, for example, whatever it is, and however it works, is an argument to the effect that the traditional view of our learning of sensation terms depends on the illusion of self-presence. Although interpreted differently by different authors, the fundamental argument is that the original item which was supposedly ostensively defined, and the item present to consciousness, are – appearances to the contrary – not logically independent of one another, and so the justification proceeds as it were by reference to itself only. Locating the 'logocentric' fallacy involved here is the cornerstone of many different anti-metaphysical philosophical schools. Some examples: deconstructionists refers to the illusion of 'self-presence'; Madhyamika Buddhists talk derogatorily of 'svabhava'; Wittgenstein writes: "'A thing is identical with itself.'" - There is no finer example of a useless proposition....' (PI §216).

mundane realm. Now, a clear effect of this Platonism is to make it seem that asking and answering a set of questions - about the 'real' motion of bodies - were a plausible project; and that, in fact, to ignore such questions would be a scientific scandal, an instance of extreme prejudice. Another effect is to make the very concept of motion (and for that matter, space and time as well) unintelligible - at least in as much as it is suggested that the idea of motion relative to a frame is a conceptual derivative of, is secondary to, the idea of absolute motion.

What this achieves is to inaugurate a fruitless program of research. It looks profound and deep, but this is because an epistemic reading is given of the metaphysical difficulties encountered. In Newton's case, various *complex procedures* for determining the absolute motion of objects, motion which is itself in no way detectable by the senses, are bolstered by a metaphysical sleight of hand. (Newton supposes - in fact, *stipulates* (and here is the give-away) - that those 'fixed' stars which do not move relative to one another are 'absolutely' still.) In this way, the epistemological program is handed over to the physical scientist.

I said in the previous paragraph that the program of research was fruitless. But this is perhaps an exaggeration, in as much as scientists, the explicit goal of whom is to work out the absolute motion of this or that, will most likely along the way discover all sorts of more or less interesting facts about the cosmos. Nevertheless, because these facts are unlikely to fit clearly into any over-arching theory, the one which initiated their collection being metaphysically spurious, they are unlikely to contribute to our *understanding* in any helpful way. They will not have the consequences they are thought to have, and are more than likely to simply build up in piles in unread papers and textbooks, their apparent significance to one generation help captive by one 'picture' of the subject matter being quickly forgotten by the next generation in the grip of another such 'picture'. That this state of affairs might be true of much work in the human sciences, and that this might have something to do with the interference of conceptually suspect premises, has long been suspected.<sup>15</sup>

---

<sup>15</sup> I hope that the systematic account given here of the origins and structures of the theoretical confusions will do something to aid situate and ground such critiques. Classic critiques have been provided by Wittgenstein, Polanyi and others. For specific examples in psychology, see Ilham Dilman's *Science and Psychology*.

The picture just painted is bleak, and in fact too bleak. For thankfully not all, or even most, of the initiative for diverse research projects comes from the metaphysical source discussed. In Newton's case, and contrary to his own declaration, hardly any of it did. A more realistic account of the influence of the metaphysical theory would see it, rather than as founding basis, as a *standing concern*. The standing concern being one of unwittingly being sucked into *quasi*-empirical projects, having such theoretical mischiefs infect one's research, and thereby diminish its explanatory breadth and power. Language lays the traps for us (Wittgenstein), and we are in constant danger of falling into them. To avoid falling into such traps, then, must be the aim of any critical and reflective scientific body of theory.

As expanded throughout the first half of the thesis, and applied in the second, mentalism, especially in its mechanistic form (i.e. 'cognitivism'), represents just such a standing concern to psychological theorising. Sometimes it will doubtless be found that, on re-examining the question which articulates our initial perplexity, that it was badly formed, that it imported with it confusions about the categorical structure of the discourse of the subject matter in question. In such cases description can indeed replace explanation, for the supposedly scientific explanatory project was unreal and didn't answer to our real needs. But in other cases the initial perplexity will indeed demand an empirical answer, and here the task is to articulate the question and pursue the quest for understanding whilst continually avoiding the distractions caused by unwittingly falling into 'Newton's error' along the way.

The case of absolute motion is just one of many analogies from physics that could have been chosen. (Another common (yet sillier) question asks why people living in Australia do not just 'fall off'.) The details of the analogy are not important; it has been used here merely to show how spurious problematics can be introduced into physical theorising by overlooking the background and misconstruing the categorical structure of the discourse domain. The purpose of choosing an analogue from physics is to reveal the formal structure of the arguments that follow in chapters 2 & 3 concerning psychology, and to show that it is in no way restricted to that discipline. In the following section I shall highlight what I consider some of the essential features of the background as it appears in psychology.

### **3. The Agential Background**

### i. Grammar, Necessity, and the Background

Discussing the conditions of possibility for various aspects of discourse under the rubric of the 'background' serves to highlight both a) the essential rôle that this broader structure plays in the individuation of those concepts foregrounded against it, and b) the unobviousness of this context when such concepts are merely used in first-order practices, or are only weakly interrogated. In the case of our spatio-temporal grammar, the background is not, once brought to our attention, too difficult to elucidate. Some presupposition of a frame of reference is not merely a conceptual or grammatical, but also a *necessary*, condition for coherent talk of motion or position. Unfortunately the same cannot be said of the background, or 'Background' (with a large 'B') which is the framework presupposed by the attribution to persons of properties (thoughts, feelings, actions, knowledge, expectations, repressed desires etc. – the list is endless) distinctive of them and of other agents. Discerning the structure of the Background is a far more complex matter than limning the concept of motion; but it is a key proposition of this thesis that such a background exists as much for beliefs and acts as it does for positions and motions.

In what follows key aspects of the Background will be elaborated. But it may as well be admitted that, whilst what follows later depends on the preliminary analysis given here, and whilst aspects of the thesis of Background are argued for in more detail in the next 2 chapters, there is no possibility in the space provided for a defence of the general idea of such holistic constraints on personal predication.

Let me at least expand on the above-mentioned distinction between conceptual or grammatical, and necessary, conditions. With respect to the discourse of motion, although we do not learn to use terms such as 'moves' by being presented with some particular *truth rule* for the use of the word, the normative use can nevertheless be easily formulated *as* a rule. The *conceptual or internal* relation between talk of motion and talk of a reference frame is therefore, at the same time, indicative of a *necessary* condition. Necessary conditions can be understood as indefeasible conceptual conditions. They typically describe those parts of discourse which are introduced into the language in the first place by means of rules (scientific and mathematical terms are sometimes introduced in this manner: a metre is a hundred centimetres), but they also obtain for those parts of language where we allow the dictionary to play a prescriptive and not merely descriptive rôle (a bachelor is an unmarried man).

At other times, however, whilst a rule may be given that seems to be not unhelpful as an indication of the meaning of a term, uses of the term can be found which show the rule to be non-universal. The rule-formulation clearly doesn't express a necessary truth. This, I would suggest, is the case for many psychological predicates; the descriptions of the normative conditions for their use, if presented as rules, would be inconceivably disjunctive and make mention of an indefinite array of defeating conditions. If learning the use of terms were a matter of 'internalising' such rules, the learning of personal predicates would be an *incredibly complex matter*. On the other hand, if the normativity of language is not ultimately to be captured by reference to *rules* but is rather a matter of *praxis*, of biologically and socially equipped beings learning a *technique*, *practical skill*, picking up *know-how*, the appearance of complexity is greatly reduced.<sup>16</sup>

But whatever the appropriate philosophy of language, the vastly disjunctive character of rule formulations is another reason why the Background is naturally self-camouflaging. For, especially given (what it wouldn't be unfair to describe as) analytical philosophy's obsession with necessary and sufficient conditions, the defeasibility of psychological predications is all too easily read as indicative of empirical matters rather than as a constitutive function of their meaning.

Intuitions concerning particular cases are perhaps not unlikely to conflict. But whether a particular case of a psychological predication being defeated is a matter of empirical fact or grammatical function can be assessed by asking the question: Can the falsity of such and such a predication in such and such circumstances be made *intelligible without the provision of grounds for the withdrawal of the predication in question*? If the answer is yes, if we feel that we can *comprehend* the facts adequately even without offering an explanation as to why the predication is not apt, then we are dealing with matters of empirical fact. If the answer is no, then we have to do with a conceptual condition – a condition for the application of the concept – the explanation offering a grammatical justification for the withdrawal of assent. True,

---

<sup>16</sup> A corollary of this position is that the normativity of language is not in the final analysis to be understood as a matter of its being rule-governed (unless this talk of rules is simply a shorthand for talk of normativity). It is the *praxis* that is the ultimate source of the normativity. Compare Wittgenstein's 'language-games' to real games such as chess. A move in the game of chess can be understood to be valid or invalid in so far as it is in accord with the rules. By contrast, it is the rules of a language-game that are to be assessed as valid or invalid insofar as they accord with the 'moves' in the language-game. C.f. Julia Tanney, *Playing the Rule-Following Game*.

what we 'ought' to say might not always be clear, and irresolvable disagreements may arise between competent speakers (and this may tell us important things about the character of folk-psychological discourse itself). But I can think of no better analytical procedure. And it should be clear from what has been suggested that this moderate approach to the character of conceptual or grammatical conditions, and the means by which they are established, in no way suggests a scepticism about the validity of distinguishing empirical from conceptual concerns, about the validity of *some* form or other of the so-called 'analytic/synthetic' distinction. Nor gives grounds for the assimilation of all such concerns to fact-stating discourses, or aids in the reduction of normativity to empirical regularity.

## ii. The Structure of the Background

Seven aspects of the Background are discussed below. They are intended to be typical and indicative, not exhaustive. Six indicate holistic constraints on the attribution of belief, understanding, action and intention and knowledge. These six aspects represent the conceptually constitutive structures which psychological predicates presuppose. The seventh aspect of the Background concerns empirical rather than conceptual preconditions. It refers to: those regularities in human behaviour that make it possible for us to take the intentional stance towards other humans, that make it possible to use personal predicates in the explanation of thought and action; those regularities without which there would be little point in talking of beliefs, desires, knowledge, understanding, perception and so on. These regularities, capacities and dispositions referred to are not individuated at the personal or intentional level – they are, as it were, reactions (or capacities to react) of the person *qua* organism rather than *qua* person; *movements*, or dispositions to move rather than *actions* or intentional capacities – and so this aspect is referred to as the Pre-Intentional Background.<sup>17</sup>

---

<sup>17</sup> The first six aspects correspond *roughly* with what Searle (*Intentionality*, pp.19ff.) refers to as the Network, and the seventh aspect with what he calls the Background. I could have used Searle's terminology, but whilst there is (as it seems to me) much of value in his treatment of intentionality, I oppose his biologism and causalism about the mind; furthermore, these naturalistic aspects of his philosophy are to my mind born of a *metaphysical* conception of the nature of philosophical questions and answers, a conception which too often takes philosophy's questions at face value.

i) The first feature of the Background can be introduced by way of Donald Davidson's 'constitutive ideal of rationality.'<sup>18</sup> Davidson's account of the attribution of propositional attitudes is explicitly holist: '... we cannot intelligibly attribute any propositional attitude to an agent except within the framework [or, as he also says, 'against the background'] of a viable theory of his beliefs, desires, intentions, and decisions.' Notwithstanding the arguably misguided emphasis on theory, the relevant consideration is that 'There is no assigning beliefs to a person one by one on the basis of his verbal behaviour, his choices, or other local signs, no matter how plain and evident, for we make sense of particular beliefs only as they cohere with other beliefs, with preferences, with intentions, hopes, fears, expectations and the rest. ... the content of a propositional attitude derives from its place in the pattern.'

An example: X says 'With this rain I shall have to leave the car at home and take the train to work.' We make sense of this by situating it within a range of other beliefs that the subject has: that the rains have flooded the roads, that there is no other simple road-route to work which would by-pass the floods, that the train-service is not similarly affected etc. If X didn't have such beliefs, we would call into question her original utterance as a genuine expression of her beliefs. If X corrected herself quickly, ('Oh, of course I could go by the back-road'), there will be no problem, but if she maintains the original assertion, we shall be forced to think of it as irrational, perhaps as manifesting a delusion. I said 'we make sense of this' and referred to the other beliefs of the subject, but there is no implication that either are conscious or unconscious acts or contents. There is simply a *default assumption* that the belief expressed by the utterance is located within a rational pattern, and this assumption will not need to be adjusted as long as various counterfactual conditions can be met – that the subject *would* say such-and-such *if* asked. In Daniel Dennett's terminology, we naturally take the *intentional stance*, instead of the design or physical stance, when encountering what to all appearances are creatures which have previously sustained and rewarded its application<sup>19</sup>. In Davidson's terms, we automatically apply the 'constitutive ideal of rationality' to the subject who's verbal behaviour we are confronted with. If we find

---

<sup>18</sup> C.f. *Essays on Actions and Events*, esp. pp. 221-3.

<sup>19</sup> C.f. *The Intentional Stance*.



our default assumption contradicted, if fail to discern a rational pattern amongst what appear to be the subject's expressions of beliefs, we 'simply forego the chance of treating them as persons.'<sup>20</sup>

ii) The second feature of the Background concerns the nature not of a subject's *beliefs* (wants, knowledge, etc), but of their *mastery of the concepts* used in their articulation.

Consider first the condition that Gareth Evans calls the 'generality constraint'<sup>21</sup>: a concept is not to be restricted for use on a single occasion; if someone has the capacity to think of A as s, and of B as t, then they also have the capacity to think of A as t, and B as s - as long as there is no ontological mismatch between predicate and object. P. F. Strawson makes a similar claim in *Individuals*, even more pertinent to our later considerations, when discussing the 'ownership' of experience: 'it is a necessary condition of one's ascribing states of consciousness, experiences, to oneself, in the way one does, that one should also ascribe them, or be prepared to ascribe them, to others who are not oneself.'<sup>22</sup>

An example: If I understand what a cow is, and what a pig is, then *if* I am to understand what it is to be pink, or brown, or four-legged, my understanding cannot only make itself manifest when I describe one of these animals. If I assent to the question: 'The cow is a quadruped', but fail to understand that the four-legged pig in front of me is similarly a quadruped, the taking of the original assent as a genuine agreement based on understanding what was asked, and what is the case, will be defeated.

Consider secondly the thesis of semantic contextualism: that words only have a meaning in virtue of the roles they fulfil within sentences, and that sentences only have the meanings they do within the language as a whole. My understanding of the meaning of terms is not, that is, *exhausted* by my ability to

---

<sup>20</sup> Stated this bluntly the position is extreme: an irrational person is still a person, just as an irrational belief (one not governed by the constitutive principle of rationality) is still a belief. What matters is that the irrationality in question is local and not global. Delusions, for example, are, notwithstanding the absence of rational relations, expressions of beliefs, but this possibility is contingent on the constitutive principle of rationality governing the subject's making of other utterances which use those terms found in the expressions of the delusional belief.

<sup>21</sup> *The Varieties of Reference*, p.103. Evans appears to conceive of our knowledge of a) what it is for something to be s as something independent of b) our understanding of what it is for A to be s, B to be s etc. But the thesis that I could not understand what it is for A to be s without also understanding what it is for A to not be s, or for B to be s, need depend on no such conceptual separation of the ability in question from its manifestations.

<sup>22</sup> *Individuals*, p.99.

point to red and only red objects when you say 'red', cows and only cows when you say 'cow.' If I could not yet use the terms to make propositions, ask questions, express myself, etc, or if I couldn't understand questions, expressions or propositions in which such terms figured, I would not be correctly described as knowing their full meaning. (If, for example, instead of saying 'Red?', you pointed to a blue object and asked 'Is this red?', and I then pointed to a red object, I could not properly be said to have a mastery of colour vocabulary). The idea is not, of course, that it is really only languages or sentences that have meanings. (To the contrary, it is not languages, but sentences and words that have meanings). Nor is it the case that a word only means something in the context of a sentence. (Again, it *is* individual words that have meanings). Rather: individual words only have the meanings they do in virtue of their possible uses within sentences. And our coming to understand these meanings is a function of our coming to understand the sentences in which they figure. And our coming to understand these sentences is a matter of coming to understand the roles they play, as one might put it, within language. Whether, for example, a certain sentence is being used to effect an ostensive definition or to state a proposition is something that has to be gleaned from the context of employment by the language-user.

The above two concerns (i) and ii)) are related by the constitutive and not merely expressive role that language plays in human life.<sup>23</sup> Possessing a language, on the view offered, is not simply a matter of being enabled to communicate ideas that could be attributed to a healthy subject irrespective of their linguistic capacities; language is rather a *transcendental precondition* of much thought and content. (Hence Wittgenstein's *Philosophical Investigations* §650: 'We say a dog is afraid his master will beat him; but not, he is afraid his master will beat him tomorrow. Why not?' (Also *PI* §250 & p.174.) The dog's form of life, being non-linguistic, provides too unsophisticated a Background to enable attributions of propositional attitudes directed towards particular moments in the long-term future). And as the Background preconditions for the attribution of propositional attitudes of any complexity include the possession by the subject of the language requisite for their expression, and because there are similar holistic constraints operating on the attribution of this latter knowledge of meanings, it follows that these semantic constraints have ramifications for the attribution of the propositional attitudes themselves. This is not to say that all intentional properties are rigidly held in place by a tight network of internal relations,

---

<sup>23</sup> Cf. Charles Taylor, *Heidegger, Language, and Ecology*.

so as to preclude the possibility of gradual learning of meanings, or the gradual accumulation of knowledge and belief. The conceptual conditions under inspection are not necessary conditions, but are flexible and defeasible; there are, moreover, no hard and fast rules specifying the conditions that apply.

iii) A third feature of the Background concerns not beliefs but feelings. It has already been noted that some small degree of rational inconsistency is allowed in the holding beliefs, although it can be stressed again that irrationality in the mind is necessarily a local and not a global affair. With the attribution of feelings and emotions we characteristically allow ourselves a somewhat greater latitude, but holistic constraints on affect ascription still exist. Some degree of consistency in a subject's emotional reactions is required before such reactions can be considered expressive of true feelings. Someone who's emotional state is highly variable and easy to manipulate can rightly be described as shallow. A more consistent subject, by contrast, can be ascribed true and deep feelings. It is not possible, for example, to be deeply in love with someone for five minutes: such an attitude involves amongst other constants a certain consistency over time of feelings of fondness and tenderness.

iv) A fourth aspect to the Background concerns the way in which the above mentioned attitudes are interdependent. This is obvious in the case of belief and understanding: one cannot believe something that one cannot understand. But the reverse is also true: a concept user necessarily has certain beliefs and a degree of knowledge too. Furthermore, it is essential before one can be said to know or to understand that one has attitudes or concerns, that things *matter* to one. (This is one of the reasons why understanding and knowledge is not attributable to a computer: such an object cannot be seen as a subject for it cannot sustain attributions of concerns: it does not have its *own* purposes, but only the purposes for which it was designed and used<sup>24</sup>. The other reason is that a computer is not an agent – its would-be understanding does not bottom out in praxical skills.) Things cannot matter to us unless we have emotions, sensibilities and feelings. These too, then, are essential in our understanding of belief and

---

<sup>24</sup> C.f. Peter Hacker, *Wittgenstein: Meaning and Mind*

understanding – they are not merely colourations of the psyche which interfere with rational functioning, but rather preconditions for talk of rationality in the first place.<sup>25</sup>

v) The fifth condition concerns the internal relations between intentions and desires and *actions*. Someone who wants things is necessarily an agent: they can manifest their intentions in action, they have the possibility of achieving their goals. Only if a subject is disposed to act on their desires, *ceteris paribus* (i.e. bearing in mind the possibility of defeating conditions: stronger desires and the like, and bodily injury), can they be said to have such desires in the first place. This relation between mind and action holds true for beliefs and knowledge of meanings too: a subject e.g. will only be held to understand a word if they can *use* it correctly in practice. It is not just the capacity to act on the environment that is relevant here, but also the capacity to take account of the environment and act accordingly. Perception therefore comes into play. My understanding of a word like 'red' is partly manifest in my capacity to discriminate red objects; it is for this reason that a congenitally blind person, however proficient they are at describing the use of 'red' (noting that it is a *colour* term, that it is the colour of *ripe tomatoes* and so on), is unlikely to be said to *fully* understand the term.

vi) The sixth condition concerns the reverse of the above internal relation. Which is to say, the necessary presence of intentions, and hence of knowledge, or at least beliefs, and wants, for the characterisation of action itself *under certain descriptions*. Perhaps I am walking across the room. You ask me: 'What are you doing?' and I reply 'I am going to get the milk out of the fridge.' What I say is obviously not a *prediction* of what I shall do, for the description is not falsified even if I get distracted, or if, say, I spot the milk on the table. The necessary background for the ascription of an action under a certain description to a subject is in part some particular intention on the part of the subject. Intentions are, as J. L. Austin put it, constitutive components in the 'machinery of action'.<sup>26</sup>

---

<sup>25</sup> This claim should appear less controversial to anyone who has read the work of Amelie Rorty or Charles Taylor.

<sup>26</sup> *A Plea for Excuses*.

vii) The final feature of the Background concerns pre-intentional aspects of a subject's behaviour. Unlike the above-mentioned considerations, the requirement of the Background here is not a logical, but rather a contingent, condition.<sup>27</sup> It is simply a fact that our being able to master a variety of discourses, understand a range of customs, make sense of different gestures, depends on our being neurologically structured in the right way. Unlike cats, people instinctively react to a pointing gesture by looking at the object of the gesture, not by looking at the finger. We instinctively relate to others' expressions of pain. We are naturally inclined, to borrow another example from Wittgenstein, to carry on the sequence '994, 996, 998, 1000...' with '1002, 1004', and not with '1004, 1008'. And it is because we have this natural disposition that we can master the basics of math. We withdraw from objects that burn us or cause other injury. We are able to 'read' others' facial expressions, postures and actions as expressing emotions. We can immediately understand other people's intentional actions *as* intentional actions; we do not just see movements. Certain juxtapositions and substitutions of phrases strike us as funny and we laugh. We instinctively imitate our parents when infants, and by means of such imitation we learn the rudiments of speech. Again, we instinctively react to signs of parental disapproval.

The above-mentioned saliencies are shared nearly universally amongst others in our culture, many of them with the rest of humanity. A considerable number are innate, and are essential for our development into the kind of complex beings that can sustain the application of the intentional stance. Others are learnt (humour and morality are culturally *shaped*, the capacity to relate to ink marks as text is learnt). Brain damage can have permanent effects on our Background capacities: we may be unable to recognise facial expressions, we may lose our moral sensibilities, or become unable to occupy the intentional stance. Of those grounding linguistic dispositions that are learnt, some may depend upon mastery of other aspects of language which act as a structural template for that which is being acquired. The concepts of mind for example and the possible constructions made using such concepts may be largely learnt along the lines of

---

<sup>27</sup> I include it partly because other philosophers, notably John Searle and Charles Taylor, have spelt out their own notions of Background in terms of preintentional capacities. But also because it is the disintegration of certain aspects of just this preintentional Background that results in the characteristic phenomenology of various mental disorders. Taylor restricts his investigation to the preintentional intelligibility-conferring context for *experience*. Searle claims reductively that background capacities, in which he includes our ability to speak, are 'really' neurological causal capacities of our brain. (This conflates personal capacities with sub-personal functions – *my brain* cannot speak as it doesn't have a mouth (...but luckily *I* do)).

the metaphor of the mind as an inner place. Many other such 'metaphors we live by' are considered by Lakoff and Johnston in their book of that name; the relevance of such grounding metaphorical structures to philosophical psychopathology will be considered in chapter 7.

#### **4. On What is to Follow**

The above themes get worked out in the following chapters in a variety of ways, some of which can be anticipated here. Part 2 defines and then examines the failures of cognitivism, and through this further elaborates the preconditions of mindedness. The philosophical arguments it employs are not necessarily novel; what is distinctive is more the manner of their employment and the framing of the critique in terms of alienation. Part 3 however provides novel argument in the use it makes of the critique of cognitivism to question cognitive theories of schizophrenia. Part 4 develops a sketch of a theory of schizophrenic alienation and fragmentation which draws on the discussion of the preconditions of mindedness to locate the psychotic deficiency at the true core of the human condition.

Chapter 2 continues the theme of chapter 1, arguing against the cognitivist's view of the mind as an *inner realm*, specifically the conception of perception as *input* into this realm, and the conception of thought as *inner representation*. Such a view situates the subject within or behind the body, alienating thought and perception from their actual worldly contents. From this perspective it is not hard to see why theories of cognitive processing are required that build up the external impressions of stimuli upon the sense organs into the putative perceptual sensations inhabiting the inner realm. This retreat of the subject into an inner mind is examined further in chapter 3: here the cognitivist conception of action as output (as mere bodily movement caused by psychological processes) is criticised. Given this conception – of mind as not imminent within action but situated behind it like an internal puppet pulling the strings of the body – it is not surprising that epistemologists have felt compelled to employ arguments from analogy or develop theories that suppose our appreciation of one another's mindedness is consequent upon our employment of theories or upon empathic projection. The alienating retreat of the subject in cognitivism goes still further, however, when we consider the topic of our knowledge of our *own* minds. Here the true subject actually retreats not merely behind the body but behind the mind itself, becoming an onlooker to

the goings on in this inner realm. Taking such a stance, the cognitivist supposes that our self-ascriptions are to be considered reliable because of a reliably functioning introspective faculty.

Part 3 examines cognitive theories of schizophrenia from two points of view. From a philosophical perspective it argues that the cognitive mechanisms and faculties within which schizophrenic breakdowns are postulated are entirely mythical (ch. 5) and so cannot be the locus of the schizophrenic dysfunctions. Secondly it contends that cognitivism cannot adequately theorise psychopathology, as the subject which in its theories retreats within the body and behind the mind remains a fully *sane* subject. The impression that the theories give is of the schizophrenic as a perfectly rational (alienated) person who (ch. 6) simply makes a few *mistakes* or who (ch. 4) has difficulty in translating their *coherent* thoughts into coherent action. Finally (Part 4) the sketch of the core preconditions of mindedness in chapter 1 is taken as a sketch of the preconditions of our contact with reality and of our rational integrity, and the schizophrenic condition is theorised as a partial breakdown of such Background preconditions.

## **Part 2**

# **Understanding Mindedness**



## 1. Introduction: On the Mind as 'Inner'



The developing account has so far urged a close relationship between certain theoretical problematics and the tacit adoption of an alienated stance towards that which the problematics concern. This alienated stance, it was argued, has the dual function of both i) encouraging the *constitutive embedding background* of the phenomenon in question to be overlooked, and of also ii) generating an apparent need for an *explanatory enterprise* explaining how and why the phenomenon in question has the properties it does. So, with respect to 'Newton's error', it was argued that Newton's alienated concept of ('absolute') motion, which conceptually disengages any particular motion from its constitutive background frame of reference, is what also gives rise to the explanatory project - both the answers *and*

*the questions* - of finding those 'true' motions of objects that supposedly further explain and underlie the 'merely phenomenal' observable changes in the relative position of objects.

The first chapter also gave details of the Background to the 'motions' of the mind. That is, it outlined those features of the subject which any ascription to that subject of understanding, thought, sensation, perception, action, and so on, must presuppose. The brief of this second chapter is, then: to show how the mischaracterisation of the categorical structure of psychological discourse and the correlative overlooking of the Background by that attitude to mind that is here described as *cognitivist* leads to a spurious theoretical problematic, and that much of the theorising in this field is the product of a tacit engagement with, or an entanglement within, this problematic. In brief, cognitivism, in taking an initially alienated conception of the human mind as constitutively disengaged from the world, encourages the view that the fundamental activities of the mind represent far more substantial achievements (such as becoming in touch with the 'external' world) than is actually the case, and its own theoretical endeavours are attempts to provide substantive accounts of just such (non-)achievements.

## ii. What is Cognitivism?

Talk of 'cognitivism' requires careful circumscription here, especially as no attempt shall be made to engage with cognitivist doctrines in their most up to date and swanky forms. Nor is the assumption made that all programmes pursued within cognitive science are irremediably infected with the confusions manifest in cognitivist metaphysics and epistemology. The project of this chapter is rather to describe and criticise a seductive conception of mind, which I have called 'mechanistic mentalism' or 'cognitivism', and to argue that unwitting subscription to this conception of mind and its associated problematics is (not a precondition of but rather) a *standing concern* for the cognitive sciences. This standing concern is, in the pure cognitive sciences, one of generating theory to answer a purely spurious problematic, and in the applied sector, one of mischaracterising (in cognitivist terms) the phenomenon to be explained and developing explanations which, because they are addressed primarily to the mischaracterisation, have far less explanatory power vis-à-vis the real phenomenon than might at first appear to be the case.

So what is 'mechanistic mentalism' or 'cognitivism'? Mentalism as it is to be understood here is a view of mind as an *inner realm*, a conception of our subjectivity that models the grammar, or the *logico-categorical structure*, of folk-psychological discourse on that of physical object language, and which models our capacity to say what occupies our minds on the epistemic capacity to say what occupies the space in front of our face. Mentalism treats the mind as somehow *inside* us - perhaps inside our bodies (passively sitting behind the eyes and the ears, actively sitting behind our limbs and mouths) - as an inner space populated by inner objects, which objects are open to view to the subject who's mind is in question, but who's existence in others must be inferred from their behaviour. In its materialist form (as in 'functionalism') mentalism identifies the mind literally with the insides of the head. The plausibility of this identity is not however our concern here; the focus of our investigation will rather concern *whether it is plausible to individuate mental contents in isolation from the world*.

Cognitivism combines such a mentalist ontology with a mechanist metaphysics<sup>28</sup>, mechanism being both the doctrine that this inner mind is *causally* linked to the periphery of the body, and to the outer world beyond the body, by a variety of psychological or cognitive mechanisms or processes, and also the doctrine that our cognitive capacities (our reasoning, thinking, making of inferences etc.) are manifest in inner causal processes. By means of this distinctive combination of mechanism and mentalism, cognitivism offers a particular (and fairly ubiquitous) conception of the relation between mind and body. What is essential to consciousness can be understood by reference to what goes on within the skin and behind the sense organs: it is in this sense that the mind is to be understood as *within* the body.

To avoid misunderstanding, I wish to stress once again that 'cognitivism' is here being treated stipulatively. 'Cognitivism' does not = 'cognitive science' or 'cognitive psychology', and the critique of cognitivism is not to be taken as automatically applying to such disciplines, although the fact that this label has been chosen for the philosophical position I consider erroneous and alienated is not insignificant: the argument is that, as the examples to come will hopefully demonstrate, research programmes in cognitive psychology can become rendered conceptually and theoretically less robust when they, wittingly or unwittingly, take on board a cognitivist conception of mind.

---

<sup>28</sup> Cognitivism also involves a distinctive third and first person epistemology; these will be addressed in chapter 3.

The picture of Caesar's head (Fig 2.1) at the top of this chapter functions effectively as a caricature of cognitivism<sup>29</sup>. The mentalist thesis that the mind is somehow within the head is ably illustrated, as is the idea that mental contents are occupants of this inner realm. The depiction is of Caesar recognising an eagle: light from the eagle impinges on the retina, and from then on there are a series of mental processes finally arriving at an act of inner recognition: the word 'EAGLE' is flashed onto an internal screen. The mechanist thesis is further illustrated by the cognitive processes that are then required to transform this recognition into Caesar's utterance of the name of the bird. According to the cognitivist, it is the job of the cognitive scientist and epistemologist to give an account of these mechanisms that transform retinal impulses (or auditory stimuli etc.), or 'inputs', into the inner contents of the mind (inner experiences or inner representations), that transform mental representations from one to another, and which then produce a series of bodily 'outputs' (that is, movements, including vocalisations).

The underlying argument of this chapter is that the mentalist's conception of mind as an inner realm, as constitutively shut off from the environment, is *ab initio* an invalid one; and that once this picture is put to right, once the mind is construed as constitutively including the environment, there will be no more apparent need for a story about the supposed causal processes mediating between the inner and the outer, and as constituting our cognition. However it is, ultimately, that we are to understand the relation between the mental and the physical, it is not the case that psychological phenomena are to be located within the body, as a *consequence* of sensory stimulation or as inner initiator of bodily actions. None of the grammatical facts justify the kind of literalisation of the metaphors depicting the mind as an 'inner' realm.

The first claim to be disputed is the cognitivist's assertion that we are simply *forced* to postulate inner representations, cognitive processes and tacit knowledge if we are to understand how we can do the things we do - that such concepts are *unavoidable* and necessary. By contrast it is argued that the appearance of necessity here is something that only arises if an alienated conception of human agency and rationality has already tacitly been accepted. The second stage argues that as well as not being necessary, the classical cognitivist concepts (inner representations, tacit knowledge etc.) are not in fact

---

<sup>29</sup> From p. 132 of H. R. Maturana & F. J. Varela's *The Tree of Knowledge*.

*explanatory* of the phenomena which they are posited to explain. The third and strongest stage of the argument is that such concepts *could not* be explanatory of our most fundamental cognitive capacities.

Although the argument is general and schematic, it will often be exemplified by cases drawn from the cognitive and cognitivist literature. This has two motives: The first aim is to show that the condition diagnosed is not a straw man: 'Newton's error' really does lie at the bottom of at least some cognitive theories and does diminish their explanatory potential (or: does give rise to an appearance of an unrealistic scientific and explanatory prowess). The second aim is historically minded: As discussed in Part 1 it was, in the mid-twentieth century, not uncommon for philosophers (such as Ryle, Wittgenstein, and their followers) to cast doubt on the conception of mind embraced by what became the cognitive sciences. Since then, however, the swing has been in the opposite direction, the assumption being that the arguments of the 'linguistic' philosophers were vitiated by their supposed positivism, behaviourism, verificationism etc. - in short by a reductive conception of both meaning and mind. Philosophical arguments today tend to concern the *details* of this or that theoretical claim; the idea that there might be something more *fundamentally* wrong with certain cognitive theories is commonly dismissed as philosophically arrogant and unrealistic. In what follows I aim to return to the scene of the earlier critiques of cognitivism, which I have suggested are best seen not as *embodying* a reductive view of mind and meaning, but rather as *critical* of that particular reductive conception of meaning which reduces the logical diversity of sundry language-games to those describing physical objects and our interaction with them. In so doing, the possibility of understanding the relevance of the earlier Wittgensteinian and Ryleian critiques to contemporary theories which embed cognitivist doctrines and presuppose alienated and objectified conceptions of subjectivity and agency is made clear.

Now diagnosis of course presupposes illness, and it is the task of the rest of this chapter to argue that the cognitivist conception of the mind as an inner realm is indeed fundamentally alienated and distorted. This will be done by providing examples of cognitivist problematics: within perception, the topics of perceptual 'constancy' and the issue of *how* we perceptually recognise things will be addressed (section 2), and within cognition, the issues of *how* we reason and of *how* our thoughts gain their meanings will be discussed (section 3). The general concern of such problematics is to show *how it is* that inner mind and outer world are related. Against this theoretical view, that there is room for a psychological

*explanation* (and not just a description) of *how* we see, recognise, think, act, the natural supposition is urged that these faculties are *psychologically irreducible*, and that their mention is indicative of the *end* of psychological explanation<sup>30</sup>.

## 2. Perception as Input

### i. Helmholtzian Problematics in the Psychology of Perception

A straightforward introduction of (what I argue to be) the spurious theoretical problematic that so easily gets inserted into theorising about vision - and which then encourages the invocation of all sorts of inner acts, cognitive processes, psychological apparatus and functions of which the agent is unconscious - can be found in the following quote from Frith's *Cognitive Neuropsychology of Schizophrenia* (p.74):

Long ago, Helmholtz (1866) pointed out that each time we move our eyes, our image of the world moves across the retina. Yet the world stays still. Thus we are able to distinguish between movement on the retina due to movements in the world and movements on the retina due to our own movements. In order to achieve this, a 'corollary discharge' is sent to some monitor system at the same time as a message is sent to the eye muscles. On the basis of this message, movement of the image on the retina is expected. Compensation occurs and the image is perceived as stationary. Thus, a distinction is made between movements of images due to our own eye movements and movements that are independent of us. This distinction is achieved by monitoring intentions to make eye movements.

What is striking about the above passage are not the neurophysiological facts mentioned (efference copy sent from neurones controlling eye movements to monitoring system etc.), but rather the psychological vocabulary used to describe them. No argument is made for the relevance of such facts to psychology (although this example is used by Frith to introduce his own conception of a monitoring system for our *thoughts* and *intentions* which can break down in schizophrenia - see chapters 4 & 5 below), but the principle descriptive vocabulary is nevertheless drawn from the *personal* and not

---

<sup>30</sup> The argument is against psychological accounts of their functioning, not against neurophysiological accounts of their functioning, nor against social and neurophysiological explanations of their development.

*subpersonal* discourse domain. It is said that *we* are able to distinguish movements on the retina due to different sources, that movement on the retina is *expected*, that we *perceive* our own *retinal images*, and that we form *intentions* to make eye movements, intentions which we monitor.

On the one hand this could be construed as a charming anthropomorphisation of the brain, and no harm is done. But what is of present concern is whether the facts described license the literal imputation of cognitive processes underlying perception and mediating our awareness of our environment, or whether the apparent necessity of positing such cognitive processes is due rather to a particular (alienated) way of conceiving of the person and their epistemic predicament which the description encourages. And so at the risk of pedantry it is worth while noting the literal *falsity* of the claims made: First, *we* (in the normal run of things) have no idea what is going on our retinae: we do not perceive images there, nor harbour expectations about the movement of such images. In fact, as the retinal image is generated by light *reflected* from the retinal surface, it is not something that *could* be seen by the person who's retinae are in question<sup>31</sup>. Furthermore, we do not normally intend to move our eyes when we do (our eye movements are not intentional or unintentional but rather preintentional<sup>32</sup>).

One (uncharitable) way of making some kind of sense of such remarks (if they are not viewed simply as metaphorical) is to *situate them within the context of an alienated conception of the human subject*. So: rather than our being able to turn our gaze onto and directly see the objects and animals around us, we are *trapped inside our own bodies* and so must resort to gaining information about the world from the images that appear on our retinae. But given this predicament a whole set of challenges arise, challenges which precipitate a range of questions of a form that are often found in cognitive psychological texts. I have tabulated these below, and include the Helmholtz-inspired example at the end:

Q1 Given that the retinal image is inverted, how is it that the world appears to us to be the right way up? A1 Cognitive processes intermediate between the eye and consciousness to reinvert the retinal images.

Q2 Given that the retinal image is more-or-less two dimensional, how is it that we see the world as three dimensional?

---

<sup>31</sup> Cf John Hyman, *The Imitation of Nature*..

<sup>32</sup> That is to say, they are not normally either intended nor unintended (i.e. performed accidentally), both of which distinctions refer to actions and their effects; the *movements* nevertheless occur (but are not *actions*).

A2 Cognitive processes extract information about depth from a variety of positional and shade (etc.) clues and make use of previously stored information about the shapes of objects. These cognitive processes finally give rise to a fully three dimensional inner representation, perhaps via a 2½ dimensional 'sketch'. (Marr)

Q3 Given that, as objects approach us, the retinal images of these objects increases in size, why is it that the objects do not appear to get larger? A3 Cognitive processes extract positional information to undo the apparent size increase...

Q4 Given that there are two retinae and hence two retinal images, why is it that we do not have two inner images of the world? A4 Cognitive processes recombine the two retinal images into one inner representation of the world.

Q5 How do we recognise objects when they are visually presented at various degrees of rotation? A5 Cognitive processes rotate our inner images of such objects until they are upright at which point template matching occurs. (Shepherd)

Q6 Given that the retinal image will move both when the object moves and when the eye or head moves, why don't the objects appear to be moving more often than they do? A6 Cognitive processes undo apparent image movements that are caused by bodily movements. (Helmholtz)

It would be foolish to suggest that such questions, whilst poorly formulated, and whilst not providing genuine 'problems' which research must aim to 'solve', do not indicate real areas of neurological research. The present concern however is with the formulations given and with the way in which they seem to press a need for cognitive processes. For if we *are* stuck inside our bodies, alienated from the world, and are forced to inspect our retinal images, the questions take on a quite genuine pertinence. In the normal run of things however, given that what we see are objects and animals and not images, the above questions as formulated just will not arise. The objects of our vision are single objects and not double images, hence there is no need for processes to undo this duplicity. Given that we see objects and not images, we would not expect the objects to appear to move unless they *do* move. Given that the world we see is (by definition) three dimensional and (by definition) the right way up, there is no need for cognitive processes to perform constructive or rotating work.

Consider though that if we *were* forced to inspect our own retinal images, the questions become more pertinent: if the proximal objects of vision are our images, it becomes genuinely puzzling why we don't see two of everything, why the world doesn't appear upside-down, why objects do not frequently



loom at us, seem to zoom past and shrink or expand. If this were genuinely our predicament then we should need all the help from cognitive processes that we could get - and more.

Resisting the problematic forced on us by the perceptual situation of an alienated self does not consist solely in assembling reminders about the everyday facts - of our perceptual contact with objects. After all, a metaphysically minded psychologist may start to elaborate the alienated perspective and actually deny that the everyday situation is the way it seems, or argue that their account is more fine-grained than the everyday one, breaking up our perception of objects into subsidiary components including the perception of images. This in effect is similar to Newton's investigation of motion: everyday relative motions are broken down into their constituent absolute motions. To make the analogy, inner images could be thought of as 'absolute perceptibilia'.

Similar problems however confront the alienated perspective, problems similar to those encountered within Newton's alienated conception of motion. For just as Newton deprived himself of the necessary ingredients for talk of motion (the positional background), the cognitivist seems to have exhausted their supply of those components of the perceptual system (most particularly: the eyes) which are required for the perception of images. Thus when the question that threatens regress arises again (namely: with what do we see our retinal images?), it is hard to see what answer can be given. Other problems arise: our normal ability to see normal images - to see what they represent - usually depends upon a *prior direct acquaintance* with what is represented, or with roughly similar objects, and also upon some degree of training. But this direct acquaintance seems to be unavailable to the subject who, stuck inside their own body, can only inspect their sensory surfaces.

It is not difficult to see how the cognitivist's conception of vision quite soon generates an inner realm of *inner* images - images that exist not merely on the retinal surface, but in the inner realm of the mind. For when we move our heads, the images produced on the retinae by the reflection of light off the stationary objects in front of us moves. Yet, so the story goes, what we perceive is a *stationary image*. This image then must be one that is generated from the retinal image, by a series of cognitive processes that serve to undo all the distortions that are generated in the appearance of the object by the picking up of this appearance by the visual system. It exists further down the line - perhaps in 'consciousness'.

But further difficulties now raise their heads, including the question of how we see our inner images, which, not located even in the eye, are even more difficult to understand as perceptibilia. It might be suggested that the inner images are not what is seen but rather *constitute* our perception; this however purchases its metaphysical and epistemological innocence at the cost of any explanatory power the mention of images might be thought to have. (It is not, that is, explanatory to say that when we perceive an object we have a perception of that object.) Or it might be said that our inner images are neither things seen, nor the act of seeing. But this is to avoid talking nonsense by avoiding saying anything at all. What's more, the fundamental epistemological difficulty soon rears its head: If all we see are images on the retina or images within the mind, how do we know that there is an 'external' world 'out there' which causes such images to appear? On the theory proposed, there could be no perceptual evidence for such a world. We should be restricted to making inferences from our inner images. And sceptical questions start to raise their heads (hence so much of the history of the western philosophical tradition: empirical and transcendental idealism, phenomenism, metaphysical realism, and other attempts to meet or defuse the epistemic threat posed by scepticism), sceptical questions that ultimately make us doubtful not only of the scope of our *knowledge* but also of our very *understanding* of the issues they articulate.

Such grand issues are not however the present concern; the current inquiry is into whether the facts about vision require the positing of *cognitive* or *psychological* processes or of unconscious inferences. What has been urged in the above is that this requirement is not generated by the facts but only by an alienated conception of vision. Resisting this conception is a matter of denying that there are any *psychological* 'how?' questions to be asked concerning our perception, to be answered by psychological theories of how we see. What has been resisted is the mapping of the subpersonal mechanics of vision onto the personal level facts about perception. But none of this is to deny that a scientific account of perceptual processes is possible; only to deny that the genuine mechanics of vision requires supplementation by a corollary psychological paramechanics.

## ii. Fodor vs. Ryle on Perceptual Processes

So far the argument has been disengaged from actual philosophical texts and pursued only at a schematic level. It will of course hardly be possible to investigate every cognitive theory which argues that the positing of perceptual processes is a necessity when providing a psychological explanation of perception. In what follows I shall investigate Fodor's critique of Ryle's critique of cognitive processes underlying perceptual recognition, thereby returning the debate to that juncture at which cognitivism is largely believed to have won the day against the conception of mind developed by the linguistic philosophers.

In chapter 6 of *The Concept of Mind* [CM], Ryle aimed to expose mechanistic mentalism's thesis that perceptual recognition must be understood in terms of 'sensation + psychological processes'. Earlier chapters, it is worth remembering, argued against the view of the mind's relation to the body as that of a 'ghost in the machine'. That is, they took issue *both* with a mechanistic conception of the body (i.e. of behaviour), with the view that human actions are governed by mechanical laws (ch. 3 section 5), and *also* with the belief that intelligent, stupid, witty, logical, inventive, dull, silly, rash, or injudicious action is action which is accompanied (or which fails to be accompanied) by a parallel inner set of 'mental' events or processes (esp. ch. 2). As against the view that intelligent performance is essentially (or even typically) performance accompanied by inner cogitations or other acts<sup>33</sup>, Ryle argued that the essence of intelligence is rather to be located in the manner of the activity and also in (what we might call) the Background circumstances of the activity. Similarly with perception (ch. 7): Ryle argued (pp. 223ff.) that the difference between a visual sensation of a cat which is an instance of perception, and a visual sensation (for example an hallucination) of a cat which is not, is similarly to be elucidated in situational rather than causal terms. Such situational concerns must of course include the presence of the cat, but also refer to other reactions of the subject, and, if we are dealing with recognition and not merely perception, the knowledge of the subject and the manner in which this knowledge is deployed. There is no suggestion in the text that these background reactions are to be considered *logically necessary*

---

<sup>33</sup> It is worth noting that Ryle did not deny the existence of mental imagery, inner soliloquy or other such distinctively mental acts.

(In this sense he is not a behaviourist). Rather, he doubted that that those cognitive capacities manifest in action essentially involved such mental acts.

conditions, but there is nevertheless and also no doubt that Ryle takes himself to be providing a *conceptual analysis* of 'perception', 'seeing', and cognate terms.

Chapter 1 of *Psychological Explanation* contains Jerry Fodor's critique of Ryle, and introduces Fodor's own defence of mechanistic mentalism, specifically his argument that, contra Ryle, any understanding of perception is simply forced to posit cognitive processes. Let us look first at the critique.<sup>34</sup> Examples of (p. 16) 'questions that a psychologist might suppose to be paradigmatic of those that a theory of perception ought to be able to answer' include the questions 'How do we see robins?' and 'How do we recognise 'Lillibullero'?'<sup>35</sup> First of all, Fodor notes that 'perceive' and 'recognize' denote achievements, and suggests (on behalf of Ryle) that in 'typical cases 'it is the presence of a robin or the fact that it *is* 'Lillibullero' that the orchestra is playing that makes the difference between perceiving and recognizing, on one hand, and misperceiving and failing to recognize, on the other.' From this suggestion Fodor concludes (on behalf of Ryle) that 'it is to facts about what is on the lawn or facts about what is being played - and not to pseudo-facts about covert mental processes - that we must refer when questions about either perception or misperception arise.'

Against this suggestion Fodor quite rightly notes that psychology is not in the business of proposing necessary conditions for vision. To be sure the presence of a robin or 'Lillibullero' is a necessary condition for perception, but it is not a sufficient condition. So (p.17) 'may not a psychologist argue as follows: given a robin to be perceived, what determines whether it *is* perceived is the occurrence of certain mental events? That is, might it not be maintained that ... it is the occurrence of the relevant mental events that makes the difference between perceiving robins *inter alia* and not perceiving anything at all [?]' Curiously no evidence at all is given that Ryle holds the (absurd) view that the only relevant considerations as to whether or not X perceives/recognises or fails to perceive/recognise Y is the presence or absence of Y, and Ryle's discussion is not actually presented. The argument is given in the

---

<sup>34</sup> Other structurally similar debates from the literature could have been investigated; for example that between JFM Hunter (*On How We Talk*) and Noam Chomsky (*Rules and Representations*, ch. 2).

<sup>35</sup> Both of these are examples that this dissertation would single out as paradigmatic of questions one would want to reject by, arguing that the capacities referred to are psychologically basic, that further explanation must revert to a neurophysiological level - and that only on an alienated conception of mind would the questions as raised appear to be in order. In what follows I shall argue that Fodor subscribes to just such an alienated conception.

spirit of a rational reconstruction of what Ryle must have meant when he wrote that (p.225) 'The [analytically relevant] questions ... are not questions of the para-mechanical form 'How do we see robins?', but questions of the form, 'How do we use such descriptions as 'he saw a robin'?''. Specifically against this view (pp.17-19) Fodor argues that Ryle conflates the projects of psychological and philosophical analysis. Ryle's stipulation about the appropriate kinds of questions to be asking are, then, simply 'expressions of taste' that should not be 'treated as though they were arguments'. It would 'be absurd to suppose that psychological accounts of the functioning of perceptual mechanisms are intended to be necessarily true'. But it is 'precisely the absurd view that all relevant psychological truths about mental processes are necessary truths that we are led to if we follow Ryle's advice and substitute inquiries into how we use locutions that assert that a performance has come off for inquiries into those mechanisms whose functioning is essential to the performance. For, given the way in which philosophers use 'use', investigations of the first kind must arrive at necessary truths if they arrive at any truths at all.'

These are fairly strong claims by Fodor, and, I would argue, as erroneous as they are emphatic. To start at the end: There is quite simply nothing in Ryle's *The Concept of Mind* which would support the opinion that the conceptual investigations into the meaning of 'perceive' and cognates, which operate by means of examination of the normative use of such terms, are designed to turn up *necessary truths*<sup>36</sup>. Ryle frequently provides lists of the sorts of sorts of behaviours, dispositions, environmental conditions etc. that he considers essential to the analysis, with no mention of 'necessary conditions'. Secondly, Ryle is not suggesting that we supplant psychology with conceptual analysis, but rather that when we are trying to discern what perception is, what it means to perceive something, we would do better with an analysis that doesn't presuppose that this essence is to be sought in cognitive mechanisms occurring alongside the visual sensation, but rather with one that examines the context in which the sensations occur<sup>37</sup>. Thirdly, this context undoubtedly includes the presence or absence of the robin or tune, but Ryle's analysis, geared as it largely is to examining the behavioural dispositions and capacities of the subject, hardly

<sup>36</sup> Rather than conditions which may in *certain circumstances* be sufficient.

<sup>37</sup> Ryle is interested in what he calls the 'logical behaviour' or 'logical grammar' or 'logical geography' of the terms; these interests delimit his field of investigation, and he is simply not interested in the psychology of perception. The critique is aimed at the traditional epistemologist, and the psychological theorist only comes in for criticism in as much as they presuppose the confusions of the epistemological tradition.

restricts itself to this single environmental fact in spelling out the difference between perception and misperception.

So much for Fodor's initial criticism of Ryle. The issues however run a good deal deeper than might be supposed. So far I have with Fodor talked quite happily about 'visual sensations'; the premise of the analysis has been that perception involves 'certain sensations + X', where X might be either cognitive mechanisms (a la Fodor) or various expectations or what have you (a la Ryle). And so long as this form of analysis is pursued, it is easy to think that, whatever the failings of Fodor's specific arguments (against Ryle), there may not be a general truth underlying what he says. For why should perception not consist, as a matter of *psychological fact* rather than *conceptual analysis*, in a supplementation of (analytically specifiable) sensations with (empirically specifiable) cognitive processes? Just as with the internal mechanics of a combustion engine (the example is Fodor's, pp.20-21), these processes may not enter into an analytic description of what it is to be an engine (where considerations of function are more to the point), but their elucidation will nevertheless be essential if we want to appraise ourselves as to *how* an engine works, or as to how we see things.

But underlying Ryle's analysis is a deep disquiet that this treatment of perception in terms of 'sensation + X' is misguided - not so much because of misleading ways in which X tends to be characterised, but rather because a purely mythical notion of *sensation* has been invoked. Caveats concerning this can be found at the end (pp.240-244) of the chapter, as well as at the beginning (pp.200-201):

... I am not satisfied with this chapter. I have fallen in with the official story that perceiving involves having sensations. But this is a sophisticated use of 'sensation'. It is not the way in which we ordinarily use these words for a special family of perceptions, namely, tactual and kinaesthetic perceptions and perceptions of temperatures, as well as for localisable pains and discomforts. Seeing, hearing, tasting and smelling do not involve sensations, in this sense of the word, any more than seeing involves hearing, or than feeling a cold draught involves tasting anything. In its sophisticated use, 'sensation' seems to be a semi-physiological, semi-psychological term, the employment of which is allied with certain pseudo-scientific, Cartesian theories.

In brief, the normal use of 'sensation' is to signify the operation of the tactile and kinaesthetic senses; sensations are bodily feelings with bodily locations, and in as much as they play a role in visual

or auditory perception they usually impair the normal perceptual process. A stinging sensation in our eye, for example, may cause discomfort and perceptual disruption: it is hardly an essential component of unimpaired vision.

Talk of 'visual sensations' then, must be carefully unpacked. For in the psycho-physical genre to which Ryle refers, terms such as 'visual sensation' are carelessly bandied about, such 'sensations' supposedly being the causal upshot of the impact of reflected light on the retina. The cognitive processes to which Fodor refers are seemingly required to mediate between this retinal stimulation and the visual sensations themselves. But if no such sensations exist, there is hardly a role for the cognitive processes. In this respect, 'visual sensations' are the equivalent of 'inner images': the product of an alienated conception of mind that places the mind inside the body and which has the mind merely receiving information about the world gathered at the physical extremity of the body.

Ryle's own analysis, as he himself admits, therefore also needs according adjustment. Thankfully his own way of spelling out the problem - in terms of the normative use of such sentences as 'he saw a robin', rather than in terms of 'questions of the para-mechanical form 'How do we see robins?'' - allows an easy resolution. The conditions which Ryle elucidates as constitutive of perception are not conditions which occur alongside some sensation which might be identically present in a case of misperception. Rather, they enter into the very core of the sensation itself. That is to say, *the 'sensation' is nothing other than the perception or the misperception itself*. And in each case 'sensation' has a different logical form, provided by the various ascription conditions that Ryle elucidates.

To spell out the issues in a more modern terminology,<sup>38</sup> we must not simply assume that perception has a *non-disjunctive* analysis, that cases of perception and hallucination have some ingredient - some 'highest common factor' (such as a sensation) - and not merely some *description* - in common. The view of the mind as inner of course implies just such an analysis: if the mind cannot reach beyond the confines of the body, that which the similar descriptions of perceptions and hallucinations describe will in each case be something internal. But whilst it is undeniable that *in some sense* a perception and an hallucination are, from the subject's point of view, indistinguishable, this does not entail that both involve some identical subjective episode. For firstly, whilst the subjects may issue identical *reports* of

---

<sup>38</sup> C.f. William Child, *Causality, Interpretation and the Mind*, ch. 5 section 2.

what they saw or hallucinated, what they saw or hallucinated was not a subjective episode of any kind (but rather, some object or event). And secondly, whilst the hallucination and perception may be indistinguishable in the sense that the hallucinating subject may *believe* that they have the same experience as the perceiving subject, this doxastic relation is again not a relation between subjective episodes (and furthermore need not necessarily obtain). Such observations pave the way for a 'disjunctive' analysis which typically goes hand in hand with non-mentalist conceptions of mind: the mind not being confined within the body can reach right out, in acts of perception, to embrace the objects themselves. Because such objects themselves form the content of the perceptual act, there is no need for inner representations of such object, nor therefore any need for cognitive processes to mediate in the generation of such inner representations.

### iii. Fodor vs. Ryle on Perception Recipes

A little later in *CM* chapter 7 (pp.226ff.) Ryle considers the ascription conditions for (not just perception but) perceptual *recognition*. He notes that to recognise the tune 'Lillibullero' it must be the case that not only is Lillibullero being played and heard and heeded, but also that the listener must have come upon the tune before, and also have learned it and not forgotten it (i.e. *know* the tune). He then asks what it is to know a tune, and in the analysis rather curiously (but doubtless correctly) refers us back to the capacity to recognise the tune<sup>39</sup>. In spelling out what it is for someone to recognise a tune, Ryle gives us a list of defeasible criteria (p.226):

... he will be said to recognise it when he hears it, if he does any, some or all of the following things: if, after hearing a bar or two, he expects those bars to follow which do follow; if he does not erroneously expect the previous bars to be repeated; if he detects omissions or errors in the performance; if, after the music has been switched off for a few moments, he expects it to resume about where it does resume; if, when several people are whistling different tunes, he can pick out who is whistling this tune; if he can beat time correctly; if he can accompany it by whistling or humming it in time and tune, and so on indefinitely.

<sup>39</sup> i.e. the internal relation between recognising and knowing a tune is two-way.



And he goes on to give a summary of this motley set of defeasible criteria by making use of the notion of a 'perception recipe' (p.227):

In short, he is now recognising or following the tune, if, knowing how it goes, he is now using that knowledge; and he uses that knowledge not just by hearing the tune, but by hearing it in a special frame of mind, the frame of mind of being ready to hear both what he is now hearing and what he will hear, or would be about to hear, if the pianist continues playing it and is playing it correctly. He knows how it goes and he now hears the notes as the progress of that tune. He hears them according to the recipe of the tune, in the sense that what he hears is what he is listening for.

In case this talk of recipes sounds like a concession to the intellectualist conception of performance Ryle is keen to criticise, he adds (pp.227-228):

Yet the complexity of this description of him as both hearing the notes, as they come, and listening for, or being ready for, the notes that do, and the notes that should, come does not imply that he is going through a series of operations. He need not, for example, be coupling with his hearing of the notes any silent or murmured prose-moves, or 'subsuming' what he hears 'under the concept of the tune'. ... It is not true that a person following a familiar tune need be thinking thoughts such that there must be an answer to the question, 'What thoughts has he been thinking?' or even 'What general concepts has he been applying?' What is true is that he must have been in some degree vigilant, and the notes that he heard must have fallen as he expected them to fall, or shocked him by not doing so. He was neither merely listening, as one might listen to an unfamiliar air, nor yet was he necessarily coupling his listening with some other process; he was just listening according to the recipe.

Now Fodor notes all the above, but argues as before that Ryle is only able to offer such a non-theoretical, simple-sounding, account because he *ignores* - rather than *undermines* - the interests of the traditional epistemologist and psychologist. He further argues that to answer such questions we must indeed make a genuine concession to the intellectualist conception of activity. This is what he says (pp.24-29):

Suppose we admit it to be a logical truth that someone who recognizes a rendition of 'Lillibullero' must be entertaining certain expectations, and suppose we pretend for the moment that this analysis of recognizing into sensing and expecting can be made general. It would certainly appear to be reasonable to request that such an

account say what, precisely, the relevant expectations are. To put it differently, it is reasonable to address to a theory that identifies recognizing a tune with hearing it according to a recipe the request that it publish the recipe.

But no sooner is that request taken seriously than all the classical arguments for conceptualism come trooping back. Consider, in particular, the 'recipe' for hearing 'Lillibullero' ...

It is clear, in the first place, that the set of events that one is capable of easily recognizing as a performance of 'Lillibullero' need not have any distinguishing acoustic characteristics. ... For one can recognize the tune when it is played on a warped record, transposed, played as a waltz, played as a march, and so on and on. It is important to bear in mind that, from a strictly acoustical point of view, the capacity to identify the tune in these various guises amounts to an enormous but highly specific tolerance of distortion. ...

Any serious attempt to construct a viable psychological theory of perception would have to account for this sort of tolerance; that is, it would have to account for the fact that training often generalizes to objects that may be only quite abstractly related to the trained object. I cannot imagine how this is to be done unless it is assumed that one's 'recipe for recognizing' shapes, tunes, and faces ... includes a representation of the formal structure of each of these domains and that the act of recognition involves the application of such information to the integration of current sensory inputs. ....

In short, if what the various ways of performing 'Lillibullero' have in common is something abstract, then it would appear to follow that the system of expectations that constitutes one's recipe for hearing the song must be abstract in the same sense. ....

[On] that analysis [of recognizing into sensing and expecting], the relevant expectations must be complex and abstract on the ground that perceptual identities are often surprisingly independent of the existence of physical uniformities among stimuli. Since it is precisely in order to explain this perceptual 'constancy' that psychologists and epistemologists have traditionally supposed that unconscious inferences and other paramechanical transactions will be needed, it seems relevant to remark that Ryle's treatment has begged all the issues that such constancy raises.

And, according to Fodor, Ryle himself pretty much exposes the shortcomings of his own account (pp. 27-28):

Ryle very nearly gives the show away when he mentions 'expecting those bars to follow which do follow ... beating time correctly ... etc.' as among the performances that would indicate that a tune has been recognized. For of course one *can* give some account of the information the hearer must employ in recognizing a tune, if one allows oneself such notions as 'bar,' 'note,' 'measure,' and 'tempo.' This is hardly surprising since such notions have developed precisely in the context of attempts to provide a vocabulary that is abstract enough to represent the common features of acoustically different renditions of a tune. ... Since however these musicological concepts *are*

abstract ... To admit that they are required to describe the perceptual recipes for tunes is simply to admit that learning to recognize tunes involves internalizing and applying complex concepts - presumably as the result of correspondingly complex mental operations.

Once again I shall work through Fodor's arguments in a reverse order. Let it be agreed that the concepts needed to abstractly describe both the tune and the expectations constitutive of recognising the tune will need to be abstract - that is, involve such notions as 'bar', 'measure' etc. This in itself in no way entails that someone who can recognise the tune, someone who harbours the expectations in question, has mastered the relevant concepts. A three year old may be able to sing along to one of their favourite records, but need not have mastered any of the concepts which would be required to describe the tune. Describing the tune musicologically is a quite different ability than being able to sing along to it. Furthermore, introduce a variation into the record and the child may well be surprised, a surprise that might best be (extensionally) described in terms of an expectation that the quaver B would be followed by a C sharp and not a D, even though the child may possess none of these concepts.<sup>40</sup> (Consider also: tying our shoelaces, riding a bike, walking along. The description of such procedures may be too complex to be mastered by someone who could nevertheless engage in them perfectly.)

To translate, now, these thoughts into Ryle's idiom: if the 'recipe' for the knowledge in question has an essentially formal description, then hearing a series of notes 'according to the recipe' (as Ryle says - pp.227 & 228) for the tune need not involve 'knowing the recipe' (as Fodor says, p.27. Ryle never talks about 'knowing the recipe'). As Ryle remarks (p.227): 'He hears [the notes] according to the recipe for the tune, *in the sense that what he hears is what he is listening for.*' And there is no suggestion in what Ryle writes that someone who is expecting this >>> note, which happens to be a middle C, need know that the note is a middle C, nor any other such (musicological) facts about it.

This point is closely related to another. Fodor argues that if what the various renditions of Lillibullero have in common is something abstract then 'it would appear to follow that the system of expectations that constitutes one's recipe for hearing the song must be abstract in the same sense.' But

---

<sup>40</sup> In *Does Cognitive Psychology Rest on a Mistake?* John Heil argues this point forcefully against Fodor's intellectualist account of how we learn a language (p.325): 'Fodor's mistake is to confuse the mechanics of description with the doings of persons engaging in the activities which the description purports to describe.'

this is simply an assumption. Even if there were a formal requirement to describe someone's expectations in such a way that covers both my expectation at time  $t_2$  that note G will come next in rendition 1, and my expectation at time  $t_2$  that note A will come next in rendition 2 (we could describe this as my expectation that the *subdominant* or *supertonic* or what have you will occur at  $t_2$ ) - and it is not clear from where this formal requirement originates - this does not entail that the person whose expectations are being described needs to know either the particular (G or A) or the abstract (subdominant) description of such notes. To say then that the expectations must themselves be abstract, if this means that the person who has such expectations must be said to possess abstract concepts or 'information' or 'representations', is unwarranted.

This much offers a diagnosis of the errors of Fodor's account, and shows why we are not after all required to attribute complex concepts or thoughts to the person who recognises a tune. It does not however reveal why Fodor was inclined to argue in this erroneous way, nor reveal the root motivation for talk of complex cognitive processes. This can be achieved if, as with Frith's discussion of corollary discharge in the eye, the necessary backdrop for such motivations is taken to be (the tacit adoption of) an alienated conception of mind.

On any normal account we might assume that we can recognise the tune that the orchestra play because we can actually hear them playing it. It does not matter too much how they play it, at what speed or in what key; so long as they are still playing the tune we shall be able to recognise it. What we hear is the tune itself. If, on the other hand, we have gotten stuck inside our own bodies again, we shall not be able to hear the tune that the orchestra play, but rather have to reconstruct what is being played from the impacts on our sense organs, which now form an interface between the inner us and the outer world. The natural description of what occurs at the sense organs is in terms of acoustics, not in terms of phrases and tunes. The ears, that is, react mechanically to various acoustical stimuli; it is people by contrast that hear tunes. The project then becomes one of arriving at a description of what is played - what *tune* is played - from a description of the acoustic properties of what is heard. Naturally if anyone were to do this it should require a great deal of musicological skill, and the mastery of complex musicological concepts. And if it were to be done unconsciously it would require a great number of psychological processes to effect the various transformations. But what is not obligatory is the concept of the person as located

inside the body; it follows therefore that the idea that this person has to perform laborious reconstructive work simply in order to be able to enjoy their natural sense functions is also not obligatory.<sup>41</sup>

### 3. Thought as Representation

#### i. Mental Models and Psychological Explanation

In what has gone so far I have argued that we are not obliged to postulate mental processes in order to psychologically understand perception, not because we can psychologically explain perception without them, but because this particular quest for psychological *explanation* is born only from an alienated perspective on human mindedness. In what follows I shall first generalise this critique of unconscious processes to include those posited to explain our rational thought (which is commonly thought to be even more ripe for psychological explanation than perception) - to suggest that we are not *obliged* to think in terms of unconscious reasoning processes - before going on to suggest that it is in fact *incoherent* to attempt such an account when what is to be explained are our fundamental rational capacities.

P. N. Johnson-Laird [J-L]'s book *Mental Models* [MM] is a popular and classic text of contemporary cognitive psychology. In some respects it attempts to move away from the intellectualist tradition (the positing of unconscious formal reasoning processes) encouraged by the mentalist conception of mind. J-L argues that formal inferences can often be replaced, in the psychological explanations of cognitive performance, by informal 'mental models'. The nature of the *explanans* is not however my concern; what shall be argued is that, whether we are considering mental models or formal inferences, the explanatory task that they are invoked to perform is unreal.

Consider how *MM* begins (pp.1-2):

---

<sup>41</sup> A corollary point, not argued here, is that the world in which the alienated subject lives has become (as John McDowell writes) *disenchanted*. The form of naturalism on offer, which defines the natural order solely in terms of physical properties, leaves no logical space in the world for such things as *tunes*. The real world is steel grey, and its colours and tastes, flavours, values and meanings are from some other order altogether. (McDowell's aim in *Mind and World* is to provide a conception of the natural order that includes meaning (etc.) within the world, with which a human mind can come into direct contact - with no need, then, for the kind of reconstructive work (on 'sensory inputs') envisaged by the cognitivist.)

Suppose, for instance, that you are told two things about a group of people in a room:

Some of the children have balloons

Everyone with a balloon has a party hat

And you have to formulate a conclusion that necessarily follows from them. Like most people, you should have little difficulty in drawing a valid conclusion, but do you have introspective access to how you did so? ^ The conclusion is undoubtedly obvious, very much more obvious than which of the many possible methods of deduction people actually use in drawing it. Protocols from more complicated problems are likewise silent on a number of matters, and it is these silences that betray the fact that introspections are at best glimpses of a process rather than detailed traces of its operations.

Introspection is not a direct route to understanding the mind and, as far as we know, there is no such route.

What J-L simply assumes is that if we do not consciously employ some strategy or heuristic in solving the deduction, then there must be an 'implicit' or unconscious strategy at work. We may not be aware of employing any such strategy, and indeed we often have nothing to say about how we reason, but far from taking this as evidence that there *is* nothing to say about *how* we reason – that there is no strategy employed, J-L takes this to indicate that we reason using an *unconscious* strategy.

In what follows I shall argue that there is no obligation to think that there is an answer to the psychologist's question concerning *how* we reason. But an important caveat must precede this. By way of analogy consider a seed drill. In asking *how* it is that the seeds come out of the drill, we may be inquiring either into the mechanism that delivers them to the end of the drill pipes, or into the pattern they make on exiting - to their density in the soil, for example. The first answer takes the question to be a demand for an explanation, whilst the second supposes only that a description is being requested. Nothing in what follows is designed to question whether there is a sensible question to be asked about how we reason if this is understood as a request for description (do we reason well or badly, for example; what errors are we likely to make?); what is questioned is whether there is an analogous psychological question to the request for explanation of seed drill mechanics.

Now *MM* is a treatise in empirical and not *a priori* psychology, and an evaluation of its empirical content is not attempted here. More specifically: it is possible to understand (much but not all of) the

book as providing a set of testable procedures with which our reasoning accords, and the further assumption that such procedures are actually employed in our reasoning can (often) be happily set aside. In terms of the analogy, the theories contained in the book can be compared to descriptions of how the seeds fall out of the drill, and not explanations of why they come to fall out in this way. Or in Ryle's unfortunate terminology, the book could be said to describe our reasoning recipes, and the suggestion that we tacitly know or unconsciously apply such recipes could be dropped. There is nothing invalid about such descriptions, although it could be argued that the common cognitivist tendency to dress them up as explanations is both pretentious and duplicitous, in that it makes the hypotheses presented appear to have far more explanatory prowess and theoretical interest than they actually possess. What I shall argue in the following, however, is just that *if* we take the rhetoric of mental models at face value, then there are really no grounds for positing such models.

That J-L's mental models are supposed to be construed quite robustly can be seen from the role they play within the theory, of replacing the unconsciously deployed laws of logic or formal Aristotelian rules of inference frequently posited by cognitive theorists (such as Piaget and Fodor). Instead of such rules, the implicit strategies that we supposedly employ are mental models - mental representations that enable us to come to the right conclusions without being super-logicians:

The psychological core of understanding, I shall assume, consists in your having a 'working model' of the phenomenon in your mind. If you understand inflation, a mathematical proof, the way a computer works, DNA or a divorce, then you have a mental representation that serves as a model of an entity in much the same way as, say, a clock functions as a model of the earth's rotation.

J-L reports that many cognitive scientists find implausible his renunciation of the doctrine of mental logic (p.131):

Explicit inferences based on mental models, however, do not need to make use of rules of inference, or any such formal machinery, and in this sense it is not necessary to postulate a logic in the mind. This claim, as I know from the reaction of audiences to whom it has been addressed, is both hard to understand and hard to believe - it is viewed as almost on a par with the Pelagian heresy in some quarters.

What underlies this incredulity is presumably the presumption that if someone is engaging in a rule-governed procedure, such as Johnson-Laird's topic of *making inferences*, then they must know the rules that govern the procedure in question. The capacity for engaging in normative behaviour must involve knowing the norms by which that behaviour is to be assessed. (For how else, it is questioned, can we recognise when we have strayed from correct procedure, if not by comparing our performance with the rule?)

This is what Fodor implies in his discussion of language learning:

Learning a language involves learning what the predicates of the language mean. Learning what the predicates of a language mean involves learning a determination of the extension of these predicates. Learning a determination of the extension of the predicates involves learning that they fall under certain rules (i.e., truth rules).

And to learn that the predicates fall under truth rules, the subject must be able to represent such rules to themselves. This leads Fodor into his notorious suggestion that we must already possess an innate language even before we learn our first language. (I shall not discuss this here.) But in any case, the general structure of the cognitivist's case goes something like this:

To say that a speaker, S, has learned a language, L, is to say that S has learned the rules that govern L. To say that S has learned the rules of L is to say that S has 'internalised' a set of formulae, R, which constitute the rules of L. The simplest way to characterise the process of 'internalisation' is to suppose that in internalising the rules of L, S comes to represent R to himself.

This is John Heil's characterisation of the situation<sup>42</sup>. And Heil suggests that the reasoning underlying this argument is simply fallacious. What I do may be *describable* in a certain way, but this in itself tells us nothing as to whether I have actually *learned the description*. And this in turn is what makes J-L's denial of the doctrine of mental logic plausible: I may correctly reason from 'Some of the children have balloons; Everyone with a balloon has a party hat' to 'Some of the children have party hats', but this

---

<sup>42</sup> Op cit. p. 325.



doesn't mean that I have 'internalised' and 'represented' to myself the rules of inference, only that my reply can be described in terms of them.

When I drew the above conclusion about the children with party hats, I reasoned to myself thus: 'If everyone with a balloon has a hat, then those children with balloons will have hats'. This however is a rare case of an almost formal inference. Consider by way of contrast the following conversation presented by J-L (p.72):

Do you have a TV set?

*Yes.*

Do you have a license for it?

*No.*

Well, it requires a license.

*Why?*

All TV sets are required by law to have a license.

J-L suggests that 'whenever an argument about a specific entity hinges on a general assertion, the chances are that its deductive form is that of a syllogism'. But even if we grant that the 'deductive form' of the above conversation is that of a syllogism, there is no ground for the suggestion that the person with the TV set comes to understand that their TV requires a license by means of a process of inference analogous to the sequence of reasoning I sketched above with respect to the hat-wearing children. I did *not* in this case reason to myself: 'All TV sets are required by law to have a license. I have a TV set. Therefore I am required to obtain a license', although I *understood*, from what was said, that I am required to obtain one.

Or consider another example (p. 94): we are presented with two premises: All the artists are beekeepers; All the beekeepers are chemists. There are various ways in which we might 'externalise the process of deduction', by which J-L appears to mean no more than: devise and implement a concrete procedure for performing the deduction. We might for example employ a group of actors to represent the different occupations. J-L writes that 'instead of arranging an external tableau, [we] could construct a mental model - an internal tableau containing elements that stand for the members of sets in just the same way that the actors did.' Well: we *might* imagine a tableau (I presume this is what 'construct ... an

internal tableau' means) to aid us in making the inference. But then again, we might not. When I drew the inference 'All the artists are chemists' I can quite honestly avow that I didn't - and *nor did I do anything else*; rather I looked at the premises and *straightway* drew the conclusion.<sup>43</sup>

There is a natural temptation to respond to the above with 'But you performed the inference *somehow!*' After all, I *performed* it!. But it is just the nature of this 'somehow' that is in question. To be sure, I drew the inference, but are there grounds for supposing that in doing so there was any *process of reasoning* going on? Johnson-Laird simply assumes, as have nearly all experimental psychologists of this century<sup>44</sup>, that thinking, understanding and reasoning are *mental processes* of one sort or another, and that like processes in general they contain stages of components.<sup>45</sup> (As processes, too, they are ripe for instantiation in neurological processes.) This, however, *is* an assumption,<sup>46</sup> and one with little by way of phenomenological or grammatical grounding: the criteria for thinking or reasoning make reference not to the existence of unknown processes but rather to the thoughtfulness or rationality of one's utterances and other actions.

The above argument does not prove that our understanding does not involve either formal rules of inference or informal mental models. What it does suggest is that grounds are required for the belief that unconscious reasoning processes underlie our everyday understanding. These grounds, furthermore, are not supplied by the kind of cursory reflection on normative practice provided by Fodor. What we say and do can often be *described* in terms of rules, but this does not mean that we have actually internalised and applied the rules in question. In Heil's terms, this would be to confuse the 'mechanics of description' with the 'doings of persons engaging in the activities which the description purports to describe'. But this

---

<sup>43</sup> C.f. Wittgenstein, *Remarks on the Foundations of Mathematics* I, §8: "The stove is smoking, so the chimney is out of order again". (And *that* is how the conclusion is drawn! Not like this: "The stove is smoking, and whenever the stove smokes the chimney is out of order; and so ...".)

<sup>44</sup> I'm thinking of (Helmholtz), Kohler, Piaget, Gregory, Bruner, as well as the cognitive scientists: Turing, Simon, etc.

<sup>45</sup> p. 40: the 'inferential mechanism'; p. 76: 'a theory of the mental machinery for syllogistic inference'; p. 80: 'the mental processes underlying syllogistic inference' etc.

<sup>46</sup> And one that has been argued against on many occasions, especially in the Wittgenstein literature; e.g. Norman Malcolm's *The Myth of Cognitive Processes and Structures*; Baker & Hacker's *Wittgenstein: Meaning and Understanding*, ch. XVI; D. Proudfoot's *On Wittgenstein on Cognitive Science*; Stuart Shanker's *Wittgenstein's Remarks on the Foundations of AI* (esp. pp. 110-120 on inferring). As well as in Ryle's *Concept of Mind*, pp. 51ff.

conclusion applies not only to inner representations of the formal rules of inference, but also to inner representations of schematic procedures for the drawing of conclusions from premises (i.e. mental models). Perhaps J-L's mental models more accurately describe how people reason. But, to return to the above seed-drill analogy, the 'how' in question is comparable to how the seeds fall on the ground, not how they come to be released from the drill in the first place. In short, there is nothing in J-L's argument to persuade us that mental models possess any psychological reality whatsoever, and nothing to show that an explanation has been given *or is required in the first place* to explain psychologically what makes it possible for us to reason and exercise our understanding.

## ii. Tacit Knowledge

Whilst we may not aptly be said to be working with syllogisms and be reasoning our way around using the rules of logic or mental models in the way that the cognitivist suggests, it is nevertheless true that in many cases we can justify our actions, decisions and utterances by appealing to rules. In fact this ability to cite adequate rules when required is *often* argued to be an important criterion for deciding whether some behaviour is or is not intrinsically normative. If I am challenged over the rationality of one of my arguments, I can appeal to the 'rules of reason' (to the Aristotelian laws for logical syllogistic reasoning) in order to justify myself. If I am playing chess and the legitimacy of one of my moves is questioned, I can justify my move by reference to the rules of the game. (The moon by contrast cannot justify its orbit by citing the equation that describes it.)

It is perhaps because of this that the cognitive theorist is likely to attribute *tacit knowledge* of rules to subjects who, whilst seemingly *playing* by the rules, i.e. acting normatively or reasoning adequately, are unable to self-ascribe knowledge of the rules in question. This tacit knowledge would mark the difference between someone who could speak and listen with understanding, or reason correctly, and someone who merely acted as if they could, someone who merely acted in accordance with the rules of right reason<sup>47</sup>. In what follows I shall argue that even though this motivation can be appreciated, it is in fact incoherent

---

<sup>47</sup> It similarly allows us to distinguish between someone making a mistake in reasoning and someone not reasoning, between someone following a rule incorrectly and someone not following a rule at all.

to attempt to explain someone's fundamental rational or linguistic activity in terms of their knowledge of the rules of right reasoning.<sup>48</sup> I shall then argue that the same incoherence attaches to J-L's attempt to explain inference-making in terms of the application of mental models.

The cognitive theorist supposes that it is possible to explain my ability to act rationally by attributing to me internalised rules of logic. Similarly when considering the ability to speak, write and understand what is said and written: the cognitive theorist is apt to attempt to explain such semantic *know-how* by ascribing to the subject tacit *knowledge* of meaning theorems or semantic rules. But if such explanations were possible it would need to true that those abilities which knowledge of the rules in question are supposed to explain are not themselves required to be able to attribute to the subject an understanding of the rules. This however is not the case.

If there is an issue about how I am able to understand some perfectly mundane sentence, the same issue will arise – yet even more so – with respect to my ability to understand the meaning theorems which provide the truth conditions for the constituent terms of this sentence. Similarly, if there is a question about how I am able to act rationally in some situation, the same question will arise – again, yet more so – with regard my capacity to correctly apply the rules of right reason. This point generalises whenever we are considering fundamental rational and linguistic capacities: *the very same capacities will always be presupposed by my ability to understand whatever rules or schema knowledge of which the cognitivist deems essential to my ability to exercise the capacities in question.*

Perhaps it be suggested that I should not need to *understand* the rules or theorems that I tacitly know, that is should be enough so long as I actually do know them. But this will not do, for what the cognitivist is after is some sort of *psychological* explanation, and what use to me are rules that I do not understand? Let it be granted then that the understanding is of a different sort to that encountered in everyday understanding (such as is commonly constitutively made manifest in those capacities which are clearly not in play in the present case<sup>49</sup> – my capacities to explain or paraphrase etc. what it is that I understand). Nevertheless, the same regress problem arises unless it can be shown that the capacities made possible by the understanding are not also required for the understanding. In however attenuated a sense, I must

---

<sup>48</sup> The argument in what follows is inspired by Julia Tanney's *Playing the Rule-Following Game*.

<sup>49</sup> Unless the subject is a professional linguist or logician.

know how to apply the rule in question, but this will require (again of course in an attenuated sense) me to understand the rule. It cannot be enough that this understanding is made manifest simply in my being able to act in accord with it. If that were the case, we are back where we started. What was wanted (by the cognitivist that is) was an explanation of *how* we are able to reason correctly, and all that has been supplied is a *virtus dormitiva*: we are able to do this or that if we understand the rules of reason, where the criterion for the possession of this understanding being our correct reasoning. A similar situation arises with tacit knowledge if the criteria for the possession of tacit knowledge are simply reduced to the possession of the know-how that this tacit knowledge is supposed to be explaining.

It might be thought that J-L's 'mental models' provide a way out of this regress. After all, one of the strengths of J-L's theory is that it seemingly does away with the need to posit a 'mental logic' which underlies our general reasoning capacities. In doing so the need to attribute to a competent language-speaker or inference-maker skills which would normally only be attributed to a logician or linguist can be obviated. But it seems that the regress cannot be blocked so easily: the capacities which will be required for the manipulation of mental models will be at least as sophisticated – even *a lot more sophisticated* – than the capacities the exercise of which the mental models are supposed to explain.

Consider: If J-L is right in saying that the following conversation (p.72) '(1) Do you have a TV set? *Yes*. Do you have a license for it? *No*. Well, it requires a license. *Why?* All TV sets are required by law to have a license.' contains a syllogism, then it is surely the case that the following (pp.97 ff.) 'effective procedure for syllogistic inference' *presupposes* the same capacities which are required to understand syllogisms. The effective procedure is: '1. Construct a mental model of the first premise.' In other words, construct 'an internal tableau containing elements that stand for the members of the sets' '2. Add the information in the second premise to the mental model of the first premise, taking into account the different ways in which this can be done.' '3. Frame a conclusion to express the relation, if any, between the 'end' terms that holds in all the models of the premises.'

Now this undoubtedly represents a scheme that people do employ *from time to time*, especially when confronted with premises the conclusion for which is not evident *straight-off*. It can hardly represent a general scheme for the capacity to follow any rational argument (as in that about the TV presented

above), however, for the same questions which (the cognitivist supposes) arise for the initial understanding will arise for the understanding of how to apply the model.<sup>50</sup>

Perhaps I 'construct a mental model' which contains an array of objects, some of which have pieces of paper (licenses) attached, and some of which are TVs (and all of which TVs have licenses attached), and then I 'add the information' that I possess a TV to the information that I have represented in my mental model, and then 'frame a conclusion' etc. Etc. But it could now be asked: How do I know that the TV I possess is represented by one of the TV representations in my mental model? It is tempting to argue that I know it because I know that my TV *is* a TV. But then again, what I know could be represented as an inference: I have a TV; the TV representations in my mental model represent all TVs; therefore my TV is represented in my mental model. And then question could be asked: How is *this* inference performed. And perhaps I then make a mental model etc.<sup>51</sup>

Again, this is not to deny that sometimes the question 'How did you reason that?' does not have an intelligible and substantive answer. But it does reinforce the naive and natural suggestion that if the question does have such an answer then the subject being questioned should be able to answer it. To the question: 'How did you draw the conclusion that your TV requires a license from the information that all TVs require licenses?' it might be natural to return a facetious answer, or just to restate the question ('All

---

<sup>50</sup> This too is the damaging flaw in J-L's above-cited clock-model comparison: J-L suggests that our mental models of situations model them in 'much the same way as ... a clock functions as a model of the earth's rotation.' But how do we understand how to apply the clock model? J-L's theory is supposed to explain our everyday understanding in terms of having working models of the phenomenon in question in our minds. This would entail our having a working mental model of the working physical clock model. But, then, how are we to understand our mental model of the clock? (It will not do to *stipulate* that mental models are self-interpreting, for such stipulation comes at the price of not providing a genuine *explanation* of the understanding in question.)

<sup>51</sup> The complexity of the understanding and knowledge required of the subject who makes mental models is manifest in J-L's specifications: (p.98): 'A crucial point about mental models is that the system for constructing and interpreting them must embody the knowledge that the number of entities depicted is irrelevant to any syllogistic inference that is drawn. ... In addition to the purely interpretative skills required to construct mental models, reasoners must appreciate the fundamental semantic principle underlying valid deduction: an inference is valid if and only if there is no way of interpreting the premises that is consistent with a denial of the conclusion.' (It could be suggested that this 'knowledge' and capacity to 'appreciate' is manifest not in any genuine discursive knowledge or understanding but rather in the fact that the subject acts *in accord with* the rules of deduction. If this were true however we are back where we started: someone draws an inference not if they know inference rules and act on them but if their inference accords with the rules of right reason. (This criterion will of course be defeasible by future actions and utterances.))

TVs require licenses so mine does'). The cognitivist may suggest that the naivete of this response is indicative of an unsatisfactory lack of depth, and that what they want to know are the *reasoning processes which allowed the deduction to be performed*, which reasoning processes *may not even be accessible to introspection*. But the price they pay for this misguided inquisitiveness is a vicious regress, a regress which could entail an age spent providing a never-ending answer to a single - arguably misguided - question.

There are at this point two options which could be had concerning inference making, neither of which will be congenial to the cognitivist. On the first option one could halt the regress at the first stage by accepting that there is often no answer to the question 'how was the inference drawn?', other than the one which provides the form of reasoning which the cognitivist wanted to get behind. On this view there are, that is, frequently no reasoning processes that go on during inference, inference being made solely in the language which clothed it at the very start. On the second option one could restrict the term 'inference' to those occasions in which the question 'how was the inference drawn?' does result in the return of the specification of a procedure by the inference-maker<sup>52</sup>. Such occasions are however rather scarce, are likely to occur mainly amongst logicians and not amongst the general public. The price of this way of saving the cognitive theory is to make it irrelevant to the study of our everyday rationality, thinking, and understanding. But it was just such topics that were the avowed interest of the cognitive theorist<sup>53</sup>.

---

<sup>52</sup> Another way of distinguishing genuine inferences (such as *MM* p. 72): 'Some arguments that are likely to be hard to understand are syllogisms. No argument that is used spontaneously in daily life is likely to be hard to understand. Therefore, some syllogisms are not used spontaneously in daily life.' from the more platitudinous cases of understanding (such as that involving TV license ownership) which might not be thought to involve inference would be to appeal to rough and ready judgements as to whether the reasoner could really be said to *understand* the reconstructed premises of the supposed inference unless they could correctly draw the reconstructed conclusion. The rationale behind this would be as follows: If a reasoner could not be attributed understanding of the 'premises' unless they could draw the 'conclusion', then so long as they genuinely do understand the 'premises' there would be no need for them to do any cognitive work in coming to the 'conclusion'. (This perhaps explains to some degree the inclination to return certain 'how did you understand/infer that?' with an 'I just did'.) Of course this criterion will provide no very definite measure as to what is and what isn't to count as an inference - but this perhaps is no very bad thing.

<sup>53</sup> Such as J-L; *MM* pp. ix ff.

### iii. Fundamental Diagnosis, and Cognitivist Theories of Meaning

The argument has now been made as to why we are not required to posit cognitive processes underlying perception and thought, and why in some cases<sup>54</sup> it is explanatorily pointless to do so. What has not yet been explained in detail is why the cognitivist is inclined to conceive of the mind in the way they do, which explanation shall now be given by drawing on the metaphilosophical themes introduced in the first chapter. By situating the discussion (including that of the vacuity of various psychological 'how?' questions) amongst the broader issues of alienation and the Wittgensteinian prohibition on philosophical theory, the structure of the cognitivist's claims and the errors diagnosed in them becomes clearer. To demonstrate the diagnostic utility of the critique of cognitivism in terms of alienation, representationalist theories of meaning will be briefly examined.

When considering perception, the critique in terms of alienation was fairly straightforward: The cognitivist adopted *ab initio* a conception of perception which took it to involve perceptual sensations (or 'perceptions' or 'representations') restricted to an inner world, sensations which then required to be placed in causal relation to the outer objects of perception via the causal mediation of the sense organs. The job of the epistemologist then was to develop theories which spell out the details of these causal transactions. If however this initial conception is rejected, and the fundamental description of perceptual acts is taken to make essential reference to the perceptual contents in the world, the need for the aforementioned epistemological theories simply drops away.

So, too, when considering rationality and comprehension: The cognitivist theorist starts with an alienated conception of our mindedness, especially of our rationality and our language abilities, taking such rationality to be fundamentally manifest in inner processes of right *reasoning*, and our understanding as grounded in representational knowledge of what is meant by what is said. Thus rational behaviour and linguistic know-how are viewed not as intrinsically and fundamentally rational and psychologically irreducible, but rather as outer behavioural acts requiring to be reanimated from the inner domain of cognitive processes and representational knowledge. This however and in any case gets the

---

<sup>54</sup> ... that is, in those cases where warranted assertibility conditions for ascription of the explanans are more logically complex, involving the possession of more complex capacities, than the assertibility conditions for ascription of the explanandum.



order of explanation quite topsy-turvy: representational knowledge presupposes know-how and hence cannot explain it, and inference-making presupposes our capacity to act rationally and so cannot explain it either. Re-invert this fundamental scheme for comprehending rationality, ground our rational and cognitive capacities in *praxis* and not *mentation*, re-imbue the behavioural with the mental, and the apparent logical space (and perceived need) for epistemological and psychological questions concerning *how* we reason and understand closes up. And re-situate the original phenomena – our linguistic expressions of understanding and demonstrations of expertise – in their natural Background context (the criteria for rationality being what is done and said – and what more is done and said – and *sometimes* what justification can be given for what is done and said), and their apparent requirement of purely ‘mental’ supplementation dies away.

Cognitivism addresses itself not only to questions of understanding but also to questions of intentionality, a concern manifest in the production of *representationalist theories of meaning*. But such theories seem to embody the same alienation and entanglement with ‘Newton’s error’ as cognitivist theories of perception and thought. In fact the common elaborations of such theories can easily be understood as defences against the sorts of regress threats examined above. Two of these elaborations – the computer metaphor and Fodor’s causal theory of representational content – are critically examined below; it is argued that they either merely disguise the regress, or fail to accommodate the essential properties of intentionality.

The general aim of representationalist theories of meaning is to account for the meaning of words and of sentences by positing an underlying set of representations in the minds of their speakers and writers. Or, in other words, the aim is to account for the intentionality of language in terms of the intentionality of thought. Or again: to account for word meaning in terms of speaker’s meaning. On a fairly standard theory, the intentionality of an ‘inner state’ (as beliefs, hopes, fears etc. are mentalistically to be conceived) is provided by its representational content, the form of which (i.e. qua belief, hope, fear) is determined by its functional rôle in the cognitive economy of the subject.

The natural objection to such a theory is well-known: When we consider paradigmatic representations, such as sketches, diagrams and models, and even when we consider what theorists are

inclined to call 'linguistic representations' (words and sentences), they cannot be said to be intrinsically meaningful. Rather, they have their meaning in virtue of the use to which they are put by human beings. This is illustrated well by Wittgenstein in *PI* §139: a picture of a man facing forward on an upward slope could just as well represent a man sliding backwards down a slope as it could a man climbing up the slope. Further representations may to some degree disambiguate the picture: perhaps there are three pictures in a row with the man in different positions on the slope in each.<sup>55</sup> Once again however this could represent a variety of different things: how are we to know that the sequence is to be interpreted as temporal and as temporally ordered with the past to the left and the future to the right? Or perhaps the man is to be thought of as staying still, and the sketcher as moving past him. What would in normal circumstances decide the intentional content of the sketch would be the sincere avowal of the sketcher. The sketch would represent what the sketcher intended it to represent. If however this intention is itself supposed to consist in a representation (playing a particular functional rôle), it is only fair to wonder whether the meaning of this representation is to be provided by a further intention. Hence the regress.

To counter this objection it is common to suggest that *inner* representations, as opposed to 'outer' representations, are 'self-interpreting' or intrinsically unambiguous. Stated that baldly the idea is of course opaque; it is a way of saying: 'admittedly there is a problem with my theory, but lets imagine instead that there isn't such a problem.' The problem is that *paradigmatic* representations are not self-interpreting, a problem which is not solved simply by stipulating that *inner* representations are self-interpreting: without further explanation we just don't know what to do with the notion of a self-interpreting representation. To try and spell out this idea, Fodor (amongst other cognitive theorists) first appeals to an analogy with the computer, and second develops a causal theory of representational content.

The computer analogy may seem to offer the cognitivist a way out of the above-mentioned regress. Computer programs are often written in 'languages' which are easy for the programmer to work with, but before the computer can run the program a compiler must translate it into a 'machine language'. This machine language is then directly 'intelligible' to the machine, and this because of the way the machine is actually physically built. This is how Fodor puts it in *The Language of Thought* p. 66:

---

<sup>55</sup> Cf Bruce Goldberg's illustrations on p. 61 of *Meaning and Mechanism*.

What avoids an infinite regress of compilers is the fact that the machine is *built* to use the machine language. Roughly, the machine language differs from the input/output language in that its formulae correspond directly to computationally relevant physical states and operations of the machine: The physics of the machine thus guarantees that the sequences of states and operations it runs through in the course of its computations respect the semantic constraints on formulae in its internal language.

To draw the analogy, our inner representations could be thought of as standing to our outer representations (our writings and utterances etc.) as the machine language of a computer stands to its programming language. Just as the machine language needs no further compiling, our inner representations need no further interpreting. Our inner representations can just 'run' on the hardware of the brain.

For the analogy to work, however, it would need to show how inner representations gain their *intentional content*, and this - on its own - it does not do<sup>56</sup>. And indeed, when we reflect further on the intentional content of computational states, the analogy seems to break down completely. If we imagine a simple accounting program, what a certain row of figures in a spreadsheet stands for will be given by what the user enters at the top of the column, and by how they understand what they do and what they intend in the first place. Similarly, what is represented by symbols in a program will be what the programmer or the computer-user (depending on the circumstance) intends them to mean. And if the internal states of the computer run through operations which are specifiable by the machine code, then they have intentional content only in so far as the machine code has content, which is only in so far as the programming language has intentional content, which in turn is only in so far as the programmer 'has content', i.e. has something in mind in what they are doing. Far from the machine code of a computer being an existence proof of self-interpreting representations, such representations possess one of the most derived forms of intentionality it is possible to imagine.

What is essentially wrong with the computer analogy is the fact that an *artefact* has been chosen as the analogical base. Artefacts have the significance they do in virtue of the *purposes of their makers and*

---

<sup>56</sup> The following is indebted to Tim Thornton's *Wittgenstein on Language and Thought*, section 5.2.3. Also to Button et al. *Computers, Minds and Conduct*.

*users*. Machine's seem to be good examples of purely mechanical systems which nevertheless can be thought of as 'processing information', and thereby provide an analogical base for mechanistic accounts of our thought and understanding - for *how* we think, *how* we mean what we do by what we say, *how* we speak a language (and the cognitivist's other distinctive psychological 'how' questions). They cannot however perform the required function, for they presuppose human intention and praxis, which are the very phenomena that the cognitivist is attempting to explain.

On the one hand, the traditional intuitive critic of mechanism is right: mechanism cannot capture what is essential about human subjectivity, cannot explain the presence of meaning in the world in naturalistic terms. But on the other hand, the trouble with mechanism is not that it cannot capture human subjectivity in naturalistic terms, but that it already *presupposes* human subjectivity, and so cannot function as an explanation of it. If we eliminate from the computer model its derivative (and therefore analogically illicit) intentionality, then it no longer looks like a plausible analogy; leave it in, however, and it no longer possesses explanatory power.

Supposing that the computer model can provide a template for understanding human thought is to commit a fallacy akin to Newton's error. The cognitivist's self-interpreting representations are like Newton's absolute motions: we are left without a way of understanding what is meant, and when the cognitivist or Newton attempts to spell out the ascription conditions, they tacitly presuppose what they are supposed to explain. The Newtonian is attempting to get behind everyday (i.e. relative to the earth) motions, to posit some explanation of such motions in terms of more fundamental categories. Their alienated conception of ('true') motion and time and space is what makes them suppose that some such project is required. In the process however they illegitimately smuggle in the actual preconditions for talk of motion, which preconditions their alienated model denies to be preconditions. The cognitivist attempts to get behind everyday expressions and cognitions and give them a more fundamental explanation in terms of what we might call 'absolute' representations. Their alienated conception of human cognition (as presupposed by, and not presupposing, praxis) is what makes them suppose that such a project is necessary. The computer model, however, smuggles back in the phenomena which are supposed to be being explained - smuggles them, that is, back into the explanation.

The causal theory of representational content is one way the cognitivist has of trying to get out of the above quandry. Several such causal theories exist - I shall consider just that provided by Fodor in his book *Psychosemantics*. I shall argue that his version of causalism, as for computationalism, either fails to sustain the normativity of intentional contents, or it presupposes and hence can't explain it.

Fodor starts off by providing a 'crude causal theory' of mental representation: this holds that a mental representation has the representational content it does in virtue of its being the case that it is commonly caused (presumably, Fodor means: caused to come into existence) by that which it represents. In other words, what it is for a mental representation to represent A is for it to be commonly caused by A.

This causal theory really is too crude, however, for as Fodor notes (p. 101), it seems that it makes no room for the normatively essential notion of *misrepresentation*. On the theory given, all of our thought would be veridical, something we obviously know is not the case. In fact, on the theory given, all of our thought would be *necessarily* veridical, a situation which makes it logically impossible for the subject possessing the inner representations to be mistaken, and therefore logically impossible for them to be said to *correctly* represent the situation either.

A simple answer would be to say that error is possible if sometimes representation 'A' is caused by A, but if at other times it is caused by B: the latter cases will represent errors when I think 'A' although it is not the case.

This however will not do, for the causal theory says that a representation represents whatever it is that causes it. If 'A' is caused both by As and Bs, then 'A' will represent a disjunction of A and B, a situation which again makes it impossible to account for error. (This is called (p. 102) the 'disjunction problem'.)

In order to solve the disjunction problem, Fodor (p.107) first remarks that

falsehoods are *ontologically dependent* on truths in a way that truths are not ontologically dependent on falsehoods. The mechanisms that deliver falsehoods are somehow *parasitic on* the ones that deliver truths. In consequence, you can only have false beliefs about what you can have true beliefs about.

And then, using linguistic representations as an analogy for mental representations, and considering a situation in which I am caused to utter 'horse' either by horses or by horsey-looking cows, he writes that such (pp. 107-8)

...causal relations aren't identical in their counterfactual properties. In particular, misidentifying a cow as a horse wouldn't have led me to say 'horse' *except that there was an independently a semantic relation between 'horse' tokenings and horses.* But for the fact that the word 'horse' expresses the property of *being a horse* ^ it would not have been *that* word that taking a cow to be a horse would have caused me to utter. Whereas, by contrast, since 'horse' does mean *horse*, the fact that horses cause me to say 'horse' does not depend upon there being a semantic - or indeed, any - connection between 'horse' tokenings and cows.

As Fodor argues (p. 108), this generates a necessary condition for B-caused 'A' tokens to be erroneous: 'B-caused 'A' tokenings are 'wild' only if they are asymmetrically dependent upon non-B-caused 'A' tokenings.' But to return to the issue at hand, which is the disjunction problem (p. 109):

... you don't get the asymmetric dependence of B-caused 'A' tokenings on A-caused 'A' tokenings in the case where 'A' means  $A \vee B$ . When we are considering disjunctive predicates, what we have is a symmetrical rather than asymmetrical dependence.

This is presented by Fodor as a solution to the disjunction problem, but it really only highlights the difficulties that any causal theory of meaning must face. What Fodor's argument shows is that we should not be prepared to talk of error unless the semantic relations in question already exist. But it is precisely to specify such semantic relations that the causal theory was brought in. It is only *given* the content of the representation that we can specify which causal relations are dependent and which are independent. But the causal theory has not shown us how to distinguish between disjunctions and errors. Rather than show us how to specify the content of the representation, it presupposes this content in drawing the distinction between dependent and independent semantic relations.

Once again the cognitivist, in attempting to provide a naturalistic theory of content, presupposes what they are supposed to be showing. The original argument was that talk of representations presupposed rather than explained our human cognitive capacities. The cognitivist attempted to

circumvent this by giving a purely causal account of representational content. This account, however, presupposed the very content that it was supposed to be explaining when it argued that disjunctive representations could be distinguished from non-disjunctive representations according to whether or not the semantic relations in question were symmetrically or asymmetrically dependent.

#### iv. Summing Up

To sum up this chapter: The cognitivist has attempted to provide various different ways in which an inner mind can be related to an outer world. They ask a variety of psychological 'how?' questions: how do we perceive, how do we recognise, how do we infer, and how are our thoughts connected to their objects. In all such cases a theory is developed which posits causal processes mediating between the inner and the outer worlds. But in all such cases the causal accounts fail: they either fail to sustain the normativity or the psychological character of the original phenomena, or they presuppose that which they are supposed to be explaining. This however need not be understood as a council of despair, for, as the argument has been, the psychological 'how?' questions, far from being obligatory, actually only arise within an *alienated* conception of human thought and action. Put this conception to right, reinstate praxis as primary and psychologically irreducible, reinstate the environment as the proximal content of perceptual acts and cognitive states, and the world can once again be unproblematically embraced within our thought.

In the next chapter (3) I shall argue that cognitivist accounts of how we tell what it is that we and others think and feel similarly presuppose an alienated conception of mindedness.

### Ch. 3. Reading Others, Expressing Ourselves

#### 1. Introduction

So far the argument has been that psychological 'how?' questions concerning perception and thought arise *posterior* and not *prior* to the construal of these faculties as, respectively, 'input' to an inner mental realm and 'inner process' occurring within that realm. The cognitivist's alienated conception of the self and mind, as trapped within the body (more specifically: within the head) is not part of an *answer* to the theoretical problematics concerning the relation of mind to world that they tackle, but is rather a precondition of the problematics themselves: a precondition of the *questions*.

The present chapter extends this argument to cognitivist conceptions of our knowledge of minds - both of our own and of other people's. The argument will be that it is only the adoption of an alienation conception of human agency and subjectivity which prompts such epistemological questions as: '*How* can we tell, given our observations of their behaviour, what other people think and feel?', and '*How* we do know what we ourselves think and feel?' What this chapter aims to show is that the simple answer '*We just can (or do)!*' is in fact the *right* answer, and that it is only on an alienated conception, which puts us at a remove from our own minds, or which puts the mind at a remove from the behaviour in which it is expressed, that any substantial epistemic achievement will seem to be involved in coming by the knowledge in question.

The cognitivist typically supposes that we come by knowledge of other people's minds by making inferences from their outer behaviour to what goes on in an inner world of mental states and processes, and that we come by knowledge of our own minds by looking within and discerning what is to be found. In place of these two theories, which I suggest involve an impossible, even unbearable 'epistemological loneliness'<sup>57</sup> of the subject (making us an *observer to* rather than *participant in* our own mental lives), a fatally disengaged conception of the conative, affective and cognitive faculties (which renders definitively enigmatic the people to whom we are closest), I develop a 'criteriological' conception of

---

<sup>57</sup> Bakan, *Clinical Psychology and Logic*.



*agency* (i.e. of mind in action) and an 'avowalist' conception of *subjectivity* (i.e. of our knowledge of our own minds). With respect to the latter it is argued that the most logically fundamental and grounding mode of self-ascription involves our speaking *from* rather *about* our own mental states. With respect to the former, our speaking of the motives of others is an epistemically straightforward matter of describing the teleological ends of their actions, and not a matter of making theory-driven inferences to an 'inner world' hidden behind some 'merely outer' behavioural realm.

The argumentative structure has two principle tacks. On the one hand the criteriological and avowalist conception of mind is defended against common cognitivist criticisms. On the other hand an attempt is made to expose the logical failures accompanying the 'epistemological loneliness' of the cognitivist's position. The inferentialist conception of our understanding of one another is argued to leave unspecified and unspecifiable the nature of that to which the inferences are made. The distance which cognitivism would open up between a subject and their own mind is also argued to be unbridgeable, and to initiate a fatal regress that would ever remove us from ourselves. Finally, the argument from chapter 2 concerning the possibility of psychological explanations of perception is extended to cover the case of everyday rational action; once again a negative conclusion is reached as to the requirement of such explanations, which only appear desirable - so the argument goes - when an alienated conception of mind has already been adopted.

As in the previous chapter, I have returned to that juncture at which cognitivism started to define itself against the anti-mentalist orthodoxy of the post-war period. Here the issues are clearer and simpler, and we can see the foundations and not the consequent metaphysical elaborations of the inferentialist conception of mind in action. What I shall argue is that these foundations are not as secure as the confidence of today's inferentialist orthodoxy would seem to suggest.

## **2. On Reading Others**

### **i. Wittgenstein on Other Minds**

Jerry Fodor has clearly become one of this dissertation's stalking horses, and his paper (written with Charles Chihara) on *Operationalism and Ordinary Language* serves as a helpful launching pad for discussion of the relation of mind to behaviour. The paper is written as a critique of, this time not Ryle's but rather, Wittgenstein's philosophical psychology, and aims to displace what I shall call his 'criteriological' conception with their own favoured cognitivist theory. In many ways the exposition of Wittgenstein's philosophy of mind and language is clear and adequate, but in what follows I shall argue that it fails in certain key respects, is unnecessarily uncharitable in others, and that the proposed alternative is both less secure than the authors suggest, and less secure than the criteriological conception it was designed to replace.

The first half of the paper (pp.384–404) provides an exposition of Wittgenstein's views<sup>58</sup>, which I shall briefly recapitulate, as they lay the foundation for the conception of mind and action deployed in the second part of *this* dissertation.

Fodor and Chihara (hereafter: 'Fodor') note that a principal aim of Wittgenstein's philosophical psychology is to dissolve the old philosophical 'problem of other minds' – the sceptical worry that we can never really know whether other people genuinely have minds, or whether in fact they just behave as if they do. A traditional approach to this problem has it that we allow ourselves the hypothesis that others have minds - that their actions are the product of thoughts and feelings - because we know that in our *own* case we have thoughts and feelings, and that we act on such thoughts and feelings in ways similar to that which is exhibited by other members of the species. This is the 'argument from analogy'. Suppose I see John hopping about after having stubbed his toe. I reason that if I were to have stubbed my toe and then hopped around this would be because I would have the feeling that I identify as pain. I therefore feel justified in saying that John likewise is in pain. Of course this is not supposed to be something that I perform every time I come to an understanding of other people; rather it represents a rational reconstruction of the ground of such understanding.

Against this argument Wittgenstein urged (*PI* §302 ff.) that it 'is none too easy a thing to do ... to imagine pain which *I do not feel* on the model of the pain which *I do feel*.' Since my own pain is definitively my own, since I cannot feel the pain of others nor they feel the pain from which I suffer, it is

---

<sup>58</sup> pp. 404–409 contain the critique, pp. 409–419 provide an alternative philosophical psychology.

hard to understand how any personal appreciation of my own sensations could ground my understanding of what it is that others feel.

The argument is not perhaps as clear as one might like, but fortunately Wittgenstein backs it up with an indefeasible argument. We might imagine the analogist holding that, whilst of course we are to respect the fact that my pains are definitively my own and those of others definitively theirs, we can understand how things stand with others, understand why they act in the way they do, if we suppose them to have pains which are *the same* - i.e. are qualitatively identical - as ours. But, Wittgenstein argues, our capacity to understand what is meant by 'the same' here *presupposes* and so cannot *underlie* our understanding of what it is for others to have sensations.

This is spelt out in *PI* §350 by means of an analogy: 'It is as if I were to say: "You surely know what 'It is 5 o'clock here' means; so you also know what 'It's 5 o'clock on the sun' means. It means simply that it is just the same time there as it is here when it is 5 o'clock.'" But this is wrong, for we understand 'it is just the same time there as it is here' by understanding that it is the same time when it is 5 o'clock on the earth *and* 5 o'clock on the sun. Our understanding of 'the same' *presupposes* our ability to deploy the concepts in the situation in question, and cannot underpin it. Similarly, when questioned what it means to say that someone else is in pain, the analogist cannot reply that it means that they have the same as we have on occasion, for our understanding of 'the same' here *presupposes* an understanding of what it is for someone else to be in pain.

To continue with the exposition: By contrast with the analogist who supposes that we know the meaning of 'pain' 'from our own case', Wittgenstein urges that we learn the meaning of this term *in the context of actual third-person ascriptions*. For example, my 'words for sensations', (he suggests in *PI* §256), are 'tied up with my natural expressions of sensation'. These natural expressions are not merely outer manifestations of pain or irritation or anger, but are rather partly *constitutive* of the phenomenon in question: to be irritated is, amongst other things, to be disposed to act like *this* >>>.<sup>59</sup>

---

<sup>59</sup> This can be compared with Ryle's analysis of understanding or perceptual recognition as given in the last chapter: if we can be said to recognise a tune, then we should be able to do a whole range of things - such as whistle the next few bars, name the piece, be surprised if it doesn't carry on as we expected, and so on. Similarly, to be said to be in pain is to be in a state in which *inter alia* we are disposed to yell and stomp about. Bede Rundle has since persuasively argued that more important behavioural criteria for

This metaphysical point has important epistemological consequences: these acts of ours are directly observable by third parties, and because they are partly constitutive of the phenomenon in question, the fact of X's pain, Y's anger or Z's irritation is no longer something that is to be understood as hiding behind the behaviour, as needing to be inferred from outer signs, but rather as something that can itself be directly apprehended. As Fodor puts it (pp.386–7): 'in many cases our knowledge of the mental states of some person rests upon something other than an observed empirical correlation or an analogical argument, viz. a conceptual or linguistic connection.'<sup>60</sup> It is because the sceptic or the analogist misrepresents the grammar of sensation ascription that they find themselves in their philosophical predicament of wondering how we could ever be justified in saying that another person is in pain on the basis of their behaviour.

On Wittgenstein's account of sensation grammar, my moans and groans and clutchings and hops are not *empirical evidence* for my leg pain. Rather they are *defeasible criteria* for the pain: given that I am acting in the way I am, there is no need for a justification of a description of me as suffering from leg pain. In this instance *justification would only be required if the ascription were to be withheld*; otherwise, we have a straightforward entailment. This is not to say that moans and groans are always criteria for pains: this criterial status depends on there not being any defeating circumstances<sup>61</sup>. When I am on the

---

sensations - more important than undirected expressions - are the *directed* reactions of a subject - avoidance of painful (and the seeking out of pleasurable) stimuli, the scratching of itches etc. Cf *Mind in Action* ch. 1.

<sup>60</sup> Fodor characterises this position as a form of logical behaviourism, and whilst noting that Wittgensteinians often oppose this label, and whilst feeling that nothing very much hangs on the label, still proposes the description on the basis of a comparison with C. L. Hull's writings. In response to Fodor, John Cook (in *Human Beings*) has stressed that whereas for Wittgenstein (and we might add, for Ryle too: c.f. Jennifer Hornsby's *Physicalist Thinking and Conceptions of Behaviour and Bodily Movements, Actions and Epistemology*) behaviour means fully intentional *actions*, for Hull it meant mere *movement*. If the behaviourist program is the naturalistic one of attempting to reconstruct or salvage as much as is good of mental functions out of merely physically defined motions, then both Wittgenstein and Ryle should surely be counted out.

<sup>61</sup> C.f. John McDowell's *Criteria, Defeasibility and Knowledge*. P. M. S. Hacker (in the first edition of *Insight and Illusion*) amongst others elaborated the defeasibility claim as holding that moans and groans are always criteria for pains, but that sometimes the criteria are defeated and the pain ascription does not hold good; whilst McDowell holds that moans and groans are only sometimes criteria for pains, and that whenever they are criteria for pains, the pain ascription will hold good. A discussion about the criteria for 'criteria' is not however to the point; what matters is what all the above would undoubtedly now agree on: that the criteria do not provide mere *evidence* that another is in pain, and that when someone else is in pain, and I see their pain behaviour,

stage, for example, my moans will be rightly taken as constituting my *simulation* of being in pain. We only understand that such acting is simulating *being in pain*, however, because in the normal case such behaviour *is* constitutive of my pain. The possibility of pretence *underlines rather than undermines* the constitutive character of pain behaviour<sup>62</sup>.

This defeasibility, and also the distinction between criteria and what Wittgenstein calls 'symptoms', is nicely illustrated by Fodor with an example from basketball (pp.391–393). There are sundry ways in which we can tell that a field goal has been scored in basketball: perhaps we listen for the roars of the crowd, or look out for the changing score on the score-board. These indices however are merely 'symptoms' of a goal's having been scored. In this respect they compare with an inflamed throat as a symptom of angina. A criterion of a field goal having been scored is the ball's going through the basket; and a criterion of angina is the presence of a certain bacillus in the bloodstream. The criteria are *defining features* of the event or state, whereas the symptoms only count as evidence for the goal or illness because we have reason to believe that they are reliable indicators, that they are reliably found to co-occur along with the defining feature. Nevertheless, the ball going through the basket does not always count as a goal: this event is only a criterion for a goal if it occurs within the context of a game of basketball, and no foul has been committed just beforehand, and so on. (Similarly, the presence of the bacillus is only a criterion for the illness if one is actually ill from having it). And likewise with pain and pain behaviour.

These remarks are related to the possibility of language-learning by both Wittgenstein and Fodor. We can learn that various features or events are symptoms for such and such only if we already know what the criteria for such and such are. And if we can come to understand the meaning of some term Y without learning that 'X is something on the basis of which one tells that 'Y' applies, [then] X cannot be a criterion of Y.' (Fodor p. 394). The mentalist supposes that all the behaviours by reference to which we learn the meaning of 'pain' function merely as co-occurring symptoms of pain. According to

---

what I see is *that they are in pain*. The fact of their pain is directly encounterable by me; I am not epistemologically at a remove from their supposedly 'inner' experience.

<sup>62</sup>Cf D Z Phillips, *Epistemic Practices: The Retreat from Reality*, pp. 34–35.

Wittgenstein, this would make it impossible for us to ever learn the meaning of 'pain', as we would never learn what it is that the (supposed) symptoms are symptoms *of*.

## ii. The Case of Dreaming (Fodor vs. Wittgenstein)

In his critique of what I have called Wittgenstein's 'criteriological' philosophical psychology, Fodor uses the example of *dreaming*. On Wittgenstein's analysis, someone's sincere say-so as to what they have been dreaming about is criterial for the dream content. This position involves what I shall refer to throughout this chapter as 'avowalism'. Avowalism holds that for much of our mental lives, especially that involving sensation, but also that of dreaming, what a subject sincerely says about their own sensation or dream *cannot coherently be questioned*. There is (so the argument goes) no room in the language-game of self-ascription for talk of our being *in error* about what we dream or what sensation we feel<sup>63</sup>. For this reason what we say (our *avowal*) can be treated as a criterion for how things are with us: my dream report can be treated by someone else as better than merely empirical evidence for the content of my dream.

On Wittgenstein's account, the mentalist misrepresents the grammar of mind by simply assimilating it to the grammar of physical objects. As concerns first-person epistemology, Wittgenstein argues that our inalienability from ourselves and the distinctive authority we enjoy concerning our own mental lives are underwritten by *grammar* (not a faculty of *inner sense*) and indicate a *categorical* difference between folk-psychological self-ascriptions and our judgements about other matters. As concerns the relation of mind to behaviour, Wittgenstein argues that behaviour is through-and-through mental, that our mentality is first and foremost manifest in our engaged action; once again the mentalist simply conflates categories and overlooks essential grammatical differences when they treat the mind as an inner cause of a distinct and isolable world of outer behaviour.

The difference between Wittgenstein's and the mentalist's treatment of the mind is particularly clear when considering dreaming. Dreaming is perhaps that aspect of our mental lives that is most naturally construed in a mentalist fashion - as a private inner phenomenon unrelated to our praxis and about which

---

<sup>63</sup> Avowalism is spelled out in more detail in section 3 of this chapter.

we (the dreamers) may be in error. It is probably for this reason that Wittgenstein's follower Norman Malcolm took on the challenge of providing a non-mentalist account of dreaming in his book of the same name, that debates about scepticism (especially since Descartes) have often involved the possibility that we may all along be dreaming, and that Fodor focuses on dreaming in his challenge to Wittgensteinian philosophical psychology.

The mentalist, to repeat, considers that my dream report, my sincere say-so as to what I have been dreaming, is merely a fallible expression of a judgement, a judgement which could be right or wrong depending on the accuracy of my memory of the 'inner process' which occupied me during sleep. Against this, Wittgenstein argues (*PI* pp. 222–223) that the 'question whether the dreamer's memory deceives him when he reports the dream after waking cannot arise, unless indeed we introduce a completely new criterion for the report's 'agreeing' with the dream, a criterion which gives us a concept of 'truth' as distinct from 'truthfulness' here.' Fodor provides five criticisms of Wittgenstein's position, criticisms which argue that the criteriological and avowalist accounts, far from doing justice to allegedly distinct aspects of our mindedness, in fact involves the philosophical psychologist in insuperable difficulties, difficulties which can only be avoided by an account of our mindedness which capitulates to the mentalist's program. I shall consider these objections in turn.

(1) Fodor notes that there are no criteria for first-person applications of many psychological predicates, and argues that 'Wittgenstein does not appear to present a coherent account of the behaviour of predicates whose applicability is not determined by criteria'. Unfortunately he doesn't substantiate this latter claim; in fact, although Wittgenstein presents a fairly extensive analysis of folk-psychological self-ascription<sup>64</sup>, Fodor fails either to present it or to argue against it. Furthermore, Wittgenstein's analysis is not that certain folk-psychological *predicates* (e.g. to 'dream') are not governed by criteria, but that there are no criteria of correctness for folk-psychological *self-ascriptions*. This observation vitiates a later comment (p.417) of Fodor's:

---

<sup>64</sup> See section 3.

Thus, the asymmetry between first and third person uses of “dream” discussed in Section VI [i.e. the suggestion that there are criteria of correctness for third- but not first-person dream ascriptions] need not arise since there need be no criteria for “X dreamed,” *whatever* value X takes: we do not have the special problem of characterizing the meaning of “I dreamed” since “dream” in this context means just what it means in third person contexts, viz., “a series of thoughts, images, or emotions occurring during sleep.”

Quite so: ‘dream’ means dream in any English speaker’s mouth, but the asymmetry between first and third person uses was not in the criteria (or lack of them) for ‘dream’, but in the criteria for ‘he dreamt’ and the lack of criteria for ‘I dreamt’<sup>65</sup>. On an avowalist interpretation, what I say can function for someone else as a criterion for the content of my dream – and error is possible for them when for example they mishear what I say; but for me, what I am sincerely inclined to say just *is* what I dreamed – and there is no possibility of error (though see the next paragraph for a refinement of this view).

(2) According to Fodor, ‘Wittgenstein’s view appears to entail that no sense can be made of such statements as “Jones totally forgot the dream he had last night,” since we seem to have no criteria for determining the truth of such a statement.’ This is debatable.

Firstly, although in such a situation we may have no criteria for empirically determining the truth of (i.e. for *verifying*) the statement, this doesn’t mean that there aren’t criteria for dreaming that *cannot* be used in verification. In fact there is no obvious logical connection between the concept of a criterion – a non-contingent mark of the identity of some phenomenon – and the concept of verification. (The behavioural and hence observable character of dream criteria does however make empirical room for the verification of judgements about the dreams of others.) As for pain: just because we have completely forgotten that we had a headache yesterday, this does not mean that there are no criteria for pain.

Secondly, the treatment of Wittgenstein’s views here is excessively unsympathetic. We can of course imagine a situation such that if Jones had been woken at t2 he would not report having had a dream, but that if instead he had been woken earlier at t1 he would have reported a dream. In such a case it would be right to suggest that by t2 Jones had forgotten the dream he had earlier. But once again, it should be

---

<sup>65</sup> The point is not that the *term* lacks criteria of application, but that there are no criteria of correctness for the *ascriptions* in question; this is because such ascriptions are not judgements, not the kind of utterances which are grammatically geared up for assessment in terms of correctness or incorrectness. (See section 3 below.)



noted, this relies on treating the earlier report as definitive vis-à-vis the later null report. To sustain Fodor's anti-anti-realist treatment of dreaming it would need to be considered a logical possibility that, had Jones been woken at *any* point during the night, and had he never reported a dream when questioned, he could still possibly have been dreaming. (All this is assuming, as Fodor notes, that no behavioural manifestations of the dreaming were exhibited). I submit that our intuitions about the logical possibility of this apparently dreamless sleep nevertheless being dream-full are a lot less clear, and that, when we remember the epistemological (anti-sceptical) advantages of the criteriological position, the view that dream reports are mere empirical correlates is less inviting than the Wittgensteinian alternative.

There is furthermore the not (I believe) infrequent experience of not being able to say quite what happened at some point during a dream, or of not being able to chose between two possible alternatives. (We should not be misled in this respect by the spurious degree of determinacy that secondary elaboration tends to lend to primary process experiences.<sup>66</sup>) What can seem on reflection to be a failure in memory of a dream (and thereby appear to lend support to the realist construal of dreaming) may rather indicate a constitutive indeterminacy in the dream experience itself. (If this is right then it indicates another respect in which Wittgenstein was correct to question the modelling of the grammar of dream reports on that of reports of worldly events).

(3) According to Fodor, Wittgenstein employs a counterintuitive procedure for counting concepts. But the example he gives is of a counterintuitive method supposedly employed by *Malcolm* in *his* analysis of dreams. The issues however will become important in Part 4 below, so a slight exegesis is in order, current relevance notwithstanding. In the *Blue Book*, Wittgenstein writes that 'If a man tries to obey the order "Point to your eye," he may do many different things, and there are many different criteria which he will accept for having pointed to his eye. If these criteria, as they usually do, coincide, I may use them alternately and in different combinations to show me that I have touched my eye. If they don't coincide, I shall have to distinguish between different senses of the phrase "I touch my eye" or "I move

---

<sup>66</sup> C.f. Sigmund Freud, *The Interpretation of Dreams*.

my finger towards my eye.”<sup>67</sup> According to Fodor, Malcolm follows this distinction of Wittgenstein’s and (p.407) ‘distinguishes not only different senses of the term “dream,” but also different concepts of sleep – one based upon report, one based upon nonverbal behaviour.’ Fodor then urges that this is an unnatural way of counting concepts.

What Malcolm urges (pp.62–63) is that there are some nightmares (some senses of ‘nightmare’) that are characterised by largely behavioural criteria – shouting and thrashing about, struggling, apparent fear – and that they are ‘so unlike the paradigms of normal sleep that it is at least problematic whether it should be said that [someone acting in this way] was ‘asleep’ when those struggles were going on.’ These cases of nightmare are, Malcolm suggests, the exception when we are concerned with dreams, and should not ‘obscure the fact that our primary concept of dreaming has for its criterion, not the behaviour of a sleeping person but his subsequent [linguistic] behaviour.’

First, it should be noted that Malcolm has not, as Fodor suggests, distinguished ‘different concepts of sleep – one based upon report, one based upon nonverbal behaviour.’ This distinction pertains solely to dreams, not to sleep.

Second, Malcolm does not say that there are different senses to ‘sleep’; he just urges that given the multiple criteria for assessing sleep, and given the fact that only some of them obtain during the violent kind of nightmares he imagines (we might also think of somnambulism and hypnotic trance), it is difficult to say whether or not the person experiencing the nightmare is asleep or not. The case is unlike that of the person pointing at their eye that Fodor quotes from the *Blue Book*. Rather it is like the case of understanding mooted by Wittgenstein in his *Lectures on the Foundations of Mathematics* (p. 23): ‘The use of the word ‘understand’ is based on the fact that in the vast majority of cases when we have applied certain [criterial] tests, we are able to predict that a man will use the word in question [the word that the man is supposed to understand] in certain ways. If this were not the case, there would be no point in our using the word ‘understand’ at all.’ There are various different behaviours (giving explanations of word meaning, being able to use the word correctly, paraphrasing a sentence in which a word occurs, reacting

---

<sup>67</sup> The discussion is opaque: touching something would not normally count as *pointing* to it, something characteristically done from a distance. A better example concerns *understanding* – see below.

appropriately to the utterances of others which contain that word<sup>68</sup>) all of which are criteria of understanding, and our normal use of the term presupposes that they co-occur. On occasion however they may come apart, and in such events it is unclear that there is any fact of the matter about whether or not the subject understands. (Our language, after all, evolved within the context of our *actual lives*.)

Third, Malcolm urged (p.62) not that there were different senses of dream or sleep, but that there were different senses of 'nightmare'.<sup>69</sup>

Fourth, what Malcolm really argued was that when we suggest that someone is dreaming on the basis of their behaviour whilst asleep, it is constitutively unclear whether the behaviour is to be understood as a symptom or as a criterion of the dreaming: 'our words do not fall definitely into either alternative'. It is for this reason that Malcolm says, perhaps unjustly, that our words here 'have no clear sense.'

To sum up: it isn't clear that any of these suggestions by Wittgenstein or Malcolm reveal a counterintuitive method of counting concepts. The only distinction of 'senses' that Malcolm makes is between the nightmares that i) involve (presumably) a subject's reporting a terrible dream, or perhaps (in the traditional folk accounts) terrible dreams of a horse lying on the chest or of drowning in the sea (different senses of 'mare') or of an incubus or succubus encounter, or perhaps a dream or feeling when asleep or whilst only partially awake that one's chest is being crushed and that one can't breathe, and ii) those that only involve the subject seemingly expressing anguish, anxiety, pain, fear, whilst thrashing about and bellowing whilst not being fully responsive to their environment. And vis-à-vis this considerable distinction, it does not seem unreasonable to suppose that we have here different senses to 'nightmare' and not merely different types of nightmare.

(4) Fodor's fourth counter-argument concerns the duration of dreams. 'As Malcolm points out, the language-game now played with "dream" seems to exhibit no criteria which would enable one to

---

<sup>68</sup> C.f. John Hyman, *Visual Experience and Blindsight* p.192.

<sup>69</sup> Of course, if the different senses of 'nightmare' are not both encompassed within any single straightforward conception of a dream, and if all types of nightmare are to be considered dreams, then the different senses of 'nightmare' will ramify into different senses of 'dream'.

determine the precise duration of dreams.' (p.407) But certain scientists have, apparently, claimed to be able to measure the duration of dreams in a precise way. If the Wittgensteinian position is correct, such scientists are confused; perhaps if we give a 'charitable' reading to their research, we conclude that they are using 'dream' in a new way. Fodor however finds this implausible: 'The notion that adopting any test for dreaming which arrives at features of dreams not determinable from the dream report thereby alters the concept of a dream seems to run counter to our intuitions about the goals of psychological research.'

Why this should be so is left unclear; one might have thought that the conceptual analysis was geared up to explaining *inter alia* what counts as psychological research on dreams. But in any case, it would seem that psychological research (or more probably, physiological research, examining REMs, EEGs etc.) is aimed towards measuring periods of *dreaming* (and not *dreams*) without waking the sleeping subject. Furthermore: It has been found that REMs occur mainly (but not always) when a subject is dreaming (using the normal criteria of verbal report on being woken etc. for dreaming), and the length of time someone is in REM sleep is then taken as a good indicator of the amount of time they have spent dreaming. Also, we surely *do* possess a fairly intuitive pre-scientific understanding of the length of both dreams and episodes of dreaming: they occur for that length of time during which, if the subject is woken, they would report a continually evolving dream<sup>70</sup>. If at t1 no dream report would be forthcoming, at t2 we would have a short dream reported, at t3 we would have a longer dream reported, and at t4 we would have no change, then we would intuitively say that the dream lasted from t2 to t3. So: It seems hard to imagine scientific tests that could 'improve' on the accuracy of this kind of procedure. And it seems impossible to imagine the kinds of justification that a scientist might use to explain why their tests test for the precise duration of dreams.<sup>71</sup>

---

<sup>70</sup> So Malcolm's account would seem to require modification.

<sup>71</sup> Ch. 13 is where Malcolm discusses the temporal duration and location of dreams. As he notes on page 79, 'Empirical studies of dreaming have produced the most divergent estimates of the duration of dreams, some investigators holding that dreams rarely last more than 1 or 2 seconds: others believe that it is 1 to 10 minutes. Dement and Kleitman ... think that dreams last as long as 50 minutes and that the average length is 20 minutes. These different estimates arise solely from the employment of different criteria of measurement.'

(5) The final objection moves us towards Fodor's own philosophical psychology, which shall be examined below. Fodor notes that for Wittgenstein the relationship between EEGs and dream reports is not criterial but rather evidential. (Though it would surely be more natural to think of EEGs as symptomatic of dreams and not of dream reports). Then he argues: 'The difficulty, however, is that this makes it unclear how the expectation that such a correlation must obtain could have been a rational expectation even *before* the correlation was experimentally confirmed.' (Although it could be recalled that REM sleep is also known as 'paradoxical sleep' precisely because it *wasn't* expected that dreaming sleep would have the physiological accompaniments that it did).

Once again no demonstration of the argument is given, and it is hard to know why a criteriological conception of dreaming and dream reports must fail to accommodate an expectation that, say, the same brain events that occur during daydreaming may not also occur during dreaming. Fodor says that 'One cannot have an inductive generalization over no observations; nor, in this case, was any higher level "covering law" used to infer the probability of a correlation between EEG and dream reports.' But imagine the following: certain EEG patterns have been found during waking cases of primary process phantasy (daydreams), and on this basis we expect the same EEG patterns to emerge during night dreaming. Our inductive generalisation did generalise several (waking) instances, and it isn't clear why it shouldn't provide the ground for a judgement about what is probable during sleep. Any new instance about which we make a prediction will differ in *some* respects from what has occurred before, even if only in temporal location. (Nothing that Fodor has said makes it clear why a criteriological conception should invite the 'riddle of induction' in any more problematic a form than any other philosophical psychology).

Fodor ends by suggesting that the same difficulties beset a criteriological conception of sensation, perception, intention etc. as well as dreaming. But because these difficulties have turned out to be relatively tractable, we may be entitled to a rational expectation that the criteriological analysis of perception, intention etc. will remain intact. In fact the case of dreaming represents one of the most internal (i.e. least behaviourally-indexed) of psychological phenomena. Fodor's failure to show that there is anything wrong with the criteriological conception of dreaming leaves the field wide open for

criteriological analyses of those psychological concepts that are more intuitively dispositional in character.

### iii. Fodor on *Inference to the Best Explanation*

In the remainder of the essay (pp. 409–419), Fodor introduces his own philosophical psychology, giving an alternative model for the relation between mind and behaviour. This starts with an important and correct observation: that not all forms of evidence take the form of symptoms, *if* by symptoms we mean indicators for some phenomenon that have been *observed to co-occur* with that phenomenon.<sup>72</sup> For consider the Wilson cloud-chamber: present scientific theories treat streaks of condensation within the chamber as evidence that charged particles (ions) have passed through. (The theory suggests that ions should act as centres of condensation in the supersaturated vapour within the cloud chamber). The status of these streaks as empirical evidence (and not as criteria) seems secure enough, something to which any theory of evidential relations will need to do justice, and it is equally clear that these streaks have not been observed to co-occur with the passage of the ions in question: they are the *only* ground we have for suggesting that the ions have taken such and such a trajectory.

With this modification in place, Fodor suggests that we would do best to think of the kinds of behaviours that Wittgenstein treats as criterial for mental events as not *criterial* but rather *evidential*, in the same way that the cloud-chamber traces are evidential for the passage of charged particles. So with respect to dream reports, the idea is that someone's sincere say-so that they have just had a dream provides no firmer (and no weaker) a ground for the claim that they have had a dream than does an EEG. Similarly with intention, desire and thought: what I say about my intentions stands (so they argue) to my intentions as does the streak in the cloud chamber to the passage of an ion: it is *logically* possible that I sincerely avow an intention but not have this intention, just as it is *logically* possible for streaks to form

---

<sup>72</sup> Fodor argues that Wittgenstein insisted that the only grounds we ever have for asserting that 'Y' applies on the basis of X are either criterial or symptomatic, but provides no textual support. Hans-Johann Glock, by contrast, suggests (*A Wittgenstein Dictionary* p.94) that for Wittgenstein, symptoms 'support a conclusion through *theory and induction*', also without textual support. I shall not pursue this exegetical question.

without an ion passing through. And just as with the cloud-chamber, psychological explanations are conceived of as *inferences to the best explanation*, the form of explanation being causal.

In what follows I shall provide various objections to the 'inference to the best explanation' conception of our understanding of one another, both (first) to the specifics of Fodor's account, and then to some more general features as found more widely in the literature. In doing so I do not mean to question that *on occasion* we might come to understand why others do what they do, or come to understand what it is for others to have (e.g.) intentions, by making inferences to the best (or perhaps not to the best) explanation. (There are, I suspect, a vast number of *different* ways in which we come to appreciate what others think and feel, ranging from straightforward observation of others right through to the analysis of the countertransference, and these different modes of gaining knowledge would, it seems to me, possess an equal diversity of logical forms.) What I aim to question is whether any such quasi-scientific epistemic procedure could plausibly *ground* our interpersonal attunement and understanding.

#### **a. *Explanation in Folk-Psychology***

Before considering the more general question of the viability of inference as a way of underpinning our understanding of one another, I wish to consider some of the peculiarities of Fodor's account. For in relation to the topic of dreaming, in which context Fodor presents his theory, it seems to me both that a psychologically atypical concept has been used, and also that Fodor's theory does not work for this particular example. Consider the latter objection first: Fodor's particular analysis of (p.415):

... the concept of dreaming is *inter alia* that of an inner event which takes place during a definite stretch of "real" time, which causes such involuntary behavior as moaning and murmuring in one's sleep, tossing about, etc., and which is remembered when one correctly reports a dream ...our notion of a dream is that of a mental event having various properties that are required in order to explain the characteristic features of the dream-behavior syndrome. For example, dreams occur during sleep, have duration, sometimes cause people who are sleeping to murmur or to toss, can be described in visual, auditory, or tactile terms, are sometimes remembered and sometimes not, are sometimes reported and sometimes not, sometimes prove frightening, sometimes are interrupted before they are finished, etc.

For the sake of argument I shall treat the suggestion that dreams are 'inner events' as simply equivalent to the idea that they are 'mental events'. The trouble however is that our concept of a dream does *not* seem to be that of a mental event which has the requisite properties 'to explain the dream-behaviour syndrome'. As Malcolm noted, the behaviours (tossing, murmuring) associated with dreams are fairly peripheral to the concept, and only infrequently occur during dreams in any case. Furthermore, to say that a dream is: what is remembered when one correctly reports *a dream* is to say, well, precisely nothing. If someone tells us a dream, we do not explain their report to ourselves (to anyone) by noting that this dream report is the report of a dream. The report typically comes specified as a dream report, and if your mere appreciation of my dream report is thought by you to explain it, then the concept of explanation has become so diluted that practically anything will count as an explanation, and practically any registering of a fact an inference to the best explanation. Sometimes, and of course, 'dream' may have an explanatory role (in explaining certain murmurs and groans, to someone in another room), but in the general run of things it can no more be considered an explanatory concept than can, say, 'table', which on occasion may also have an explanatory role (consider: how is that teapot being supported? - oh, by a glass table).<sup>73</sup> It is in this respect the contrast with the roles typically played by such concepts as 'negatively charged ions' that is more striking than the similarity when considering such homely concepts as 'dream'.

To turn now to the question of the typicality of 'dream' as a psychological concept: as already noted, of all psychological concepts, those concerned with fantasy are the least connected with action and behaviour. This is brought out both in the fact that 'dream' does not seem to be an *explanatory* concept vis-à-vis behaviour, but also in the fact that it may not be directly *internally related* to action concepts. Dreams do not occupy any kind of central rôle in our understanding of one another as agents, and although there are certain internal relations to broadly behavioural concepts (the dreaming subject must

---

<sup>73</sup> With respect to Fodor's other comments about dreams - that they can be reported, have duration, can be described in sensory terms, are sometimes forgotten, are sometimes frightening - these are undoubtedly generally correct, and reveal the manner in which dreams (as opposed to numbers or expectations) are categorically suited for such description. (But this says nothing as to the typical explanatory or non-explanatory role of 'dream'.)



be asleep; or, considering daydreaming: we may not be prepared to say of an active, engaged (i.e. not sitting still and looking vacuous) conversing subject that they are having a daydream), these will not be as prominent as those for concepts such as 'desire' and 'intend' which intuitively possess a far more proximal bearing on action. True, the criteriologist considers a dream report criterial for dream *content*, but this may be as far as the internal relations between action (i.e. the report) and the 'inner event' (i.e. the dream) go for what may be a fairly atypical concept in the first place.

The two above-mentioned objections may fairly be thought to cancel each other out. It may be, that is, that when we consider more typical psychological concepts, the 'inference to the best explanation' model will be found more apposite. I shall now, therefore, turn to further more general objections to that model.

One general objection can be lodged against Fodor's model: Fodor says that a dream is an inner event which *causes* a variety of behaviours, and we may suppose that he would be happy with a generalisation which considered other 'mental events' such as pains to be the causes of the related shouts and screams and avoidance behaviour. When we consider the matter empirically, however, there is little basis for this belief in the causal role of the sensation. It seems it is just as plausible that the tissue damage may cause both the physical responses and the sensation, and that the sensation therefore does not have a causal role. In fact certain studies<sup>74</sup> concerning the timing of body movements and sensations would seem to support the idea that the sensation and certain bodily responses are not directly causally related, but are rather both effects of a common cause<sup>75</sup>. But whether or not this is generally true, the fact that it is even readily *intelligible* suggests that as a conceptual analysis Fodor's theory is inapposite. There just is no *a priori* reason to think that 'howling in pain' *demand*s a parsing as 'howling caused by pain'.

A more fundamental objection concerns Fodor's general analysis of mental concepts (p. 413):

Perhaps, what we all learn in learning what such terms as "pain" and "dream" mean are not criterial connections which map these terms severally onto characteristic patterns of behavior. We may instead form complex conceptual connections which interrelate a wide variety of mental states. It is to such a conceptual system

---

<sup>74</sup> Harth, E., *Windows on the Mind*, p. 100-1.

<sup>75</sup> Rundle, *Mind in Action*, p. 20.

that we appeal when we attempt to explain someone's behavior by reference to his motives, intentions, beliefs, desires, or sensations. In other words, in learning the language, we develop a number of intricately interrelated "mental concepts" which we use in dealing with, coming to terms with, understanding, explaining, interpreting, etc., the behavior of other human beings (as well as our own). In the course of acquiring these mental concepts we develop a variety of beliefs involving them. Such beliefs result in a wide range of expectations about how people are likely to behave.

Now no criteriolgist, and certainly not Wittgenstein, would want to disagree with the suggestion that there exist 'complex conceptual connections' which interrelate a wide variety of psychological concepts. This kind of holism about the mind is constantly embraced by Wittgenstein and made manifest in his unwillingness to attribute faculties and facilities piecemeal to subjects. But the question to which, in Fodor's exposition at least, Wittgenstein's criteriological conception of mind was addressed was the venerable 'problem of other minds', and what needs to be seen is whether Fodor's replacement inferential theory has the conceptual sources to negotiate this problem.

What Fodor tells us is that mental concepts are concepts that are used in understanding, explaining, interpreting etc. the behaviour of other human beings. This is (often) true, but the problem of other minds asks us why, in some instance or other, we are *justified* in applying this vocabulary to others. How can we conceive of the minds of others when all we see is their behaviour? How can we come by the idea that other people have minds? Relatedly, if mental phenomena are essentially inner and private, how is it that we learn the meaning of the mental terminology to which Fodor refers? To be told that we *do* use the terminology in giving explanations of one another's behaviour is not to resolve this issue.

The other component in Fodor's theory is the reference to the internal relations between mental concepts. But, even granted the constitutive importance of such relations, it is hard to see what they could contribute to a trainee folk-psychologist. That is, the trainee may learn that the concept 'belief' is internally related to that of 'desire' or 'thought' or 'intention' or what have you in such and such ways (entailment, exclusion etc.), yet be none the wiser as to what any of the terms actually *mean*, or how to employ them in practice. If there is a problem about our understanding of 'belief', it is not helped by stressing the internal relation with 'knowledge', for if this term also stands for some internal state, there would presumably be an equal problem with our understanding of 'knowledge'. What is needed is some

way into this system of concepts, a way in that is provided by the criteriologist who stresses the internal relations between observable actions and unobservable psychological states, but not by the inferentialist who views this system as conceptually closed in relation to the merely 'external' behaviour of the subject. In their focus on inference the inferentialist has provided a merely epistemic account of our understanding of one another which clearly presupposes an account of what it is that the inferences are inferences to, an account which cannot be adequately specified solely in terms of conceptual relations internal to the system of mental contents.

To see better the inadequacies of a conception of mental concepts that solely tracks internal relations between the concepts it is helpful to substitute variables for the concept terms. Thus whilst it would clearly be enlightening if one didn't know what an intention was to be told that a desire that X along with knowledge that Y entails, in such and such circumstances, an intention that Z, it is clearly not so helpful to be told that an A that X along with a B that Y entails a C that Z. The entirety of the system of internal relations amongst psychological concepts could be learnt, but, even if we were to settle for the common yet rather narrow<sup>76</sup> cognitivist view that folk-psychology is fundamentally a matter of prediction and explanation, it is hard to see how a trainee folk-psychologist would ever learn how to bring the system of concepts to bear on cases of behaviour which were felt to be in need of explanation or prediction. Similarly, someone may know the various internal relations amongst the terms making up our colour discourse (which colours are 'lighter' or 'darker' than which other colours), yet be none the wiser in being able to identify colours in the real world.

#### **b. Inference in Folk-Psychology**

The principle objections to the 'inference to the best explanation' model have not however yet been touched on; these concern the appositeness of the inferentialist's belief that in relevant respects folk-psychological understanding is akin to scientific understanding - that our folk-psychology characteristically involves *inference*, involves us in making *predictions*, and, even, is fundamentally an

---

<sup>76</sup> C.f. Michel ter Hark, *Wittgenstein and Dennett on Patterns*.

*explanatory* activity. The very term 'folk-psychology', with its implication that our understanding of one another depends upon our mastery of some kind of theory, itself tends to prejudice the debate. In what follows I shall continue to use the term, but without intending to suggest that the practical knowledge manifest in our abilities to understand one another is grounded in any such representational or even theoretical knowledge. What I shall argue is that it is only on an alienated conception of mind that our folk-psychology comes to look like some kind of theory allowing us to make inferences from the 'outer' to the 'inner' of a subject in the same way that our physical theory allows us to make inferences from a streak in the cloud chamber to the passage of a charged particle.

One objection is provided (and rather surprisingly not answered) by Fodor himself (p. 412). After having given the comparison of a mental state and an observable action with the passage of an ion through a cloud chamber and the visible streak it leaves behind, Fodor suggests that

It might be replied that the above examples do not constitute counter-instances to Wittgenstein's criterion-correlation premise since Wittgenstein may have intended his principle to be applicable only in the case of ordinary language terms which, so it might seem, do not function within the framework of a theory. It is perhaps possible to have indicators that are neither criteria nor symptoms of such highly theoretical entities as electrons and positrons, but the terms used by ordinary people in everyday life are obviously (?) in a different category.

The suggestion that all merely empirical evidences must conform to the pattern of symptoms can, as already argued, be left behind, but the point that our ordinary folk-psychological terms are not part of a theory is not touched by this concession. Our ordinary discourse for ascribing dreams and intentions to others *is, intuitively*, non- or pre-theoretical. We are not, for example, ever *taught* such a theory for ascribing intentional states to subjects.

The argument does not have to be left at this intuitive level. Many different thoughts might sit behind the objection to the view that folk-psychology embodies some kind of theory<sup>77</sup>, but one clear one turns on what is known by someone who knows a theory and who can provide explanations on the basis of the theory. Someone who rationally expects an ion to leave a trace of vapour in the cloud chamber can, on the basis of their theory, tell us *why* they expect this. It is in this respect that a theory is more than a

---

<sup>77</sup> Cf Kathy Wilkes on *Folk Psychology*.

mere suggestion - it provides us with an understanding of why it is that a charged particle should leave a condensation trace, and it is because of this understanding that the trace is seen as empirical evidence for the presence of an ion. By contrast, it isn't clear that we have the capacity, on the basis of our folk-psychological know-how, to say *why* we expect someone with an intention to issue an action aimed at the intentional object. In fact it is arguable that the demand for such an explanation is not coherent (I shall argue shortly that it is indeed misguided), nor clear what would count as a plausible answer<sup>78</sup>.

Deciding the issue involves us in a philosophical territory normally dominated by the question of whether the relation of mind to action is to be considered *causal*. As is well known, the Wittgensteinian orthodoxy suggests that causal accounts of mind are misguided, whilst the mentalist majority considers causality to be an essential notion in our understanding of mind and its relation to action.<sup>79</sup> A fundamental lynchpin for the anti-causalist's account has been the 'logical connection argument' - the argument that causal relations are fundamentally external relations - relations between essentially distinct items, and that action and (e.g.) desire are not conceptual isolates, and not therefore appositely thought of as related causally. Whether or not this or some such related criterion of causality gives a fair assessment of the structure of that concept has however been questioned, although it is notable that those who reject the criterion and thereby save the respectability of causalism do not always seem to provide any other such criterion by means of which the content of the causalist claim can be assessed.<sup>80</sup>

---

<sup>78</sup>I have no intention of simply dismissing the physicalist's *philosophical* theory concerning the causal effects on bodily movements of brain states or events which, on their theory, *are* our mental states or mental events. What is clear however is that this is a philosophical theory, and we do not need to know any such theory before we can understand one another.

<sup>79</sup>In fact even some avowedly non-mentalist philosophers (such as Bill Child, Jennifer Hornsby, Donald Davidson) are causalists: here the motivation behind the causal account is not the desire to provide a link up between the mentalist's supposedly independent 'inner' and 'outer' worlds, but is rather of a piece with such authors' *physicalist* commitments. Prominent mentalist causalists include Fodor and the Churchlands; prominent non-mentalist non-causalists include Malcolm, Rundle, Peters, Wilson and Tanney. (See bibliography for references.)

<sup>80</sup>Donald Davidson is the best-known critic of causalism: on his account internal relations hold between descriptions of items, whilst external i.e. causal relations hold between the items themselves. So: just because an intention may be related to an action under a certain description does not mean that it cannot be causally related as well. By contrast I should argue that internal relations are between concepts, and that if the 'descriptions' in question are definitive of the concepts then items internally related under such essential descriptions cannot be thought of as possessing the kind of distinctness that we intuitively suppose causally related

The argument here however will not worry itself over whether the relations between mental contents and actions are causal. What is of more interest is the 'logical connection' criterion used (misguidedly or not) to assess the causal claim, and not the causal claim which it is commonly employed to negate. The argument will be that, whether or not internally related items can be causally related, the greater the externality of the relation between the items, the greater the possibility of the citing of such causal relations being genuinely *explanatory* and the greater the provision for *predictive* possibilities. And that when folk-psychology and scientific theory are considered along this continuum, it is the contrasts that are more pertinent than the commonalities.

Consider the development of Fodor's inferential conception of folk-psychology by Paul Churchland.<sup>81</sup> Churchland notes rightly that (p. 53) a 'perennial objection is that these generalisations [i.e. the generalisations which describe the structure of folk-psychological truths] do not have the character of genuine causal/explanatory laws; rather, they have some other, less empirical status (e.g. that of normative principles or rules of language or analytic truths).' (An objection I should like to register.) But he considers that 'serious difficulties' confront any such objection, first claiming that there is a degree of categorical distinction between those folk-psychological generalisations that involve sensations and feelings (pain, hunger, grief) and those that involve the so-called propositional attitudes (belief, desire, intention). Notably, the former tend to feature in causal/explanatory laws, such as (p. 53):

- A person who suffers severe bodily damage will feel pain.
- A person who suffers a sudden sharp pain will wince.
- A person denied food for any length will feel hunger.
- A hungry person's mouth will water at the smell of food.
- A person who feels overall warmth will tend to relax.
- A person who tastes a lemon will have a puckering sensation.

---

items to possess. (It is also questionable whether Davidson's causalism is well-motivated: Julia Tanney, for example, has persuasively argued that Davidson only feels the need to supplement the rationalising relations between an agent's reasons and their actions with causal relations between the same because he has an overly determinising conception of the role of such reasons in the first place.)

<sup>81</sup> *Folk Psychology and the Explanation of Human Behaviour.*

- A person who is angry will tend to be impatient.<sup>82</sup>

Propositional attitude explanations, by contrast (p. 54) 'display the explanandum event as "rational"'. But notwithstanding this important contrast, Churchland argues that the latter should also be seen as causal in character, and this because:

Whatever else humans do with the concepts for the propositional attitudes, they do use them successfully to predict the future behaviour of others. This means that, on the basis of presumed information about the current cognitive states of the relevant individuals, one can nonaccidentally predict at least some of their future behavior some of the time. But any principle that allows us to do this – that is, to predict one empirical state or event on the basis of another, logically distinct, empirical state or event – *has* to be empirical in character. And I assume it is clear that the event of my ducking my head is logically distinct both from the event of my perceiving an incoming snowball, and from the states of my desiring to avoid a collision and my belief that ducking is the best way to achieve this.

The question that must now be asked is whether i) propositional attitudes are genuinely and normally used in *prediction*, and the related issue ii) raised as to whether the logical independence that Churchland detects between my ducking and my desire is genuine.

With respect to the first I submit that we hardly ever make predictions about what other people will do. I do not mean to imply by this that, given our understanding of other people, given our knowledge of their characters, we will not characteristically entertain a variety of *expectations* as to how they are likely to react. Knowing that my mother has a fear of spiders, I expect that she will run out of the room when Harold makes his evening excursion from under the coal shuttle. But it would, I suggest, be extravagant to call this a prediction.

---

<sup>82</sup> Before drawing the comparison note that these 'humble generalisations' are *not* all 'clearly ... causal/explanatory' in character. Whilst the first example illustrates a causal truth, it is not one explanatory of behaviour. The second is not always true, and may in any case represent an empirical generalisation and not a causal statement - or we may even decide that wincing is criterial for the sharpness of pain. The fifth may have more to do with warmth than the feeling of warmth, and so may not be properly folk-psychological. With respect to the sixth - what on earth is a 'puckering sensation'? Finally, it could be argued that impatience is a constitutive component of grammar, and that we therefore have to do with a regularity in grammar rather than (as Patricia Churchland suggests, *Neurophilosophy*, p. 299) a 'regularity in nature'.

The issue does not simply turn on a question of linguistic extravagance. One clear difference between expectation and prediction is the differential requirement for *reasoning*. I fully expect the sun to rise tomorrow, and this expectation can be attributed to me whether or not I have thought about it. All that matters in this case is that I have not entertained any thoughts to the contrary. In a sense my expectation is grounded in my past experience, but this is not to say that the past experience forms some sort of tacit *reason* such as might be postulated to occur in some equally tacit reasoning. Given that the sun has always, in my experience, risen, I should need a reason for thinking that it will *not* tomorrow, but am not in need of a reason for thinking that it *should*.

An astronomer's prediction that an eclipse will happen on a certain date, by contrast, requires thought on their part. If they genuinely predict an eclipse they should be able to furnish reasons for believing that it will happen on such and such a date.

Expectations can be entertained on the basis of a knowledge of character, and 'character' is a dispositional concept. So: I know that such-and-such a glass is brittle, and expect it to break if dropped onto a hard floor. Or: I know that John is a greedy so-and-so, so I fully expect him to eat me out of house and home. Being likely to break when dropped is just what it means to be brittle; being prone to eat a lot is just what is meant by 'greed'. It is for this reason that no reasoning or prediction is required: we could not be said to understand that someone were greedy in the first place lest we had the relevant expectation. The knowledge of character does not enable me to make a prediction – rather it has an expectation constitutively built into it. Similarly, one might think, with desire: if we know that someone wants a cup of tea then, everything else being equal, we should expect them to go and make or buy one. Explanation would be required if they didn't do so, but no explanation of why people act on their desires is required: to be disposed to act in certain ways is constitutive of what it means to have the relevant desires.

Predictions however cannot be entertained solely on the basis of a knowledge of disposition. It is for this reason that it sounds pretentious to describe our future-directed understanding based on knowledge of dispositions as a matter of prediction. It is as if we are crediting ourselves with far more intellectual activity and understanding than we engage in or possess. Knowing that Tim is an irascible fellow I expect him to get into a stew when I fail to supply the work I promised; this however is no sort



of prediction: my knowledge that he is irascible is not the *basis* of my judgement but simply *of a piece* with it.

Churchland it will be recalled suggested that my actions (my ducking the snowball) are only contingently related to my desire (to avoid it). We are now in a position to evaluate this claim. As a reconstruction of what might happen in a snowball fight the example is perhaps rather unreal (I do not believe that ducking rather than dodging, say, is the best way to avoid a snowball collision - I just act in one way or another), but let us imagine that something like it is true. Imagine that I believe and want what is said of me, that there is nothing else I want more in the circumstances (to win a bet or prove a point), that I am fit, that I see the snowball approaching with plenty of warning: Is it *conceivable* that I do not try to duck? Or to put it otherwise: are we not *logically compelled* to think of defeating conditions for an attribution of desire to avoid the snowball if, along with the belief in question, ducking is not consequent? I would suggest that such an attribution is inconceivable - that the lack of action is in the circumstances itself the strongest possible defeating condition for the desire ascription. And that action and desire are, therefore, not merely empirically but rather *internally* related.

#### **iv. The Defeasibility of Behavioural Criteria**

The logical relation that I have been considering between intention or desire and action is not straightforwardly analytical, not expressed as a necessary truth. My desire for an ice-cream does not *entail* that I shall attempt to procure one; and we do not have to do with the logical behaviourist's reduction of statements about the mind to statements about behaviour. Given this lack of analyticity, it might be asked how the criteriological conception nevertheless differs from the cognitivist's inferential theory. Since, on the criteriological position, it will always be possible to specify defeating conditions for the ascription of some psychological attitude given some behaviour, or to provide reasons why some action is not forthcoming notwithstanding the propriety of an ascription of some attitude, and given that the specification of such defeating conditions and reasons is not, it is admitted, codifiable - given all of this, in what sense is a non-analytical relation between mind and action not simply empirical but rather 'conceptual'?

The simple answer is that whereas defeating conditions *could* always be provided for the ascription of (e.g.) some desire, it is only the criteriological conception that makes the recognition of this provision a *necessity* for anyone to be considered as understanding the propositional attitudes. The conceptual relation between desire and action is manifest in the logical *requirement* that some defeating condition or other be specified if a propositional attitude ascription is to be sustained in the face of behaviour that to all appearances fails to make that attitude manifest. This simple feature makes sense too of the lack of a requirement to be able to specify some 'complete' list of defeating conditions in an analytic reconstruction of a propositional attitude concept. Whilst, in a 'strict sense', in a mode of thought governed by sensitivity to our heritage's philosophical attention to the *logically necessary* (and concomitant lack of attention to the real conceptual workings of natural languages<sup>83</sup>), a proposition about a desire does not *entail* one about an action, we may nevertheless specify as a conceptual truth that a desire will issue in action so long as some actually specifiable defeating circumstance is unforthcoming. The (cognitivist) inferentialist by contrast makes no such obligation on those who would be counted as having mastery of propositional attitude discourse. On the account Fodor offers it makes perfect sense to attribute a desire to someone who fails to act on it notwithstanding an absence of reasons for not so doing. It is against this seeming lack of rational constraint in the negotiation of the social domain that the criteriological position takes a stand.

## v. The Psychological Character of Behaviour

---

<sup>83</sup>I am thinking here especially of the idea that Wittgenstein criticised but which still finds unquestioning support: that the normativity of language - the correctness or incorrectness of uses of words - is to be understood in terms of whether such uses are or are not sanctioned by *rules* for the use of terms. (One unfortunate implication of his rhetoric of 'language-games' is that, in assimilating moves in language to the moves in a game, we are inclined to think of the appropriateness of such linguistic moves as assessable in terms of rules in the same way that moves in chess or basketball might be.) Necessary truths represent those corners of language that do conform to this ideal: they are linguistic norms codifiable as unimpeachable rules. There is no reason - nothing in our childhood experience of language learning for example - to think that the semantics of most of our language (excluding specialised extensions in scientific and logical theory) conforms to anything like this model. The rules we formulate are normally *descriptions of* and not *prescriptions for* language use, and derive their normativity from the intrinsic normativity of the language games (not vice versa).

The general pitch of the objection developed here has been that the inferentialist, in supposing that our understanding of the minds of others is based in inference or in a theory that furthermore issues in predictions, has tacitly subscribed to an alienated conception of mind. The mind has been conceived of in abstraction from those actions in which, it has here been argued, it is constitutively and not simply contingently manifest. It is because of this (supposed) logical gap between the mental and the behavioural that anything so formal as inference is supposed to be required. But it is not the case that such inferences or predictions are normally invisible to us simply because we are so immersed within our folk-psychological practice, nor because they are unconscious or tacit. Rather, they are not required, because our immersed perspective does not put us in the epistemic predicament that the inferentialist supposes.

It is not however simply an alienated conception of the propositional attitudes that sets up this epistemic divide. What the inferentialist typically supposes is that mental phenomena are commonly posited in order for us to *make sense of behaviour*, as if we could understand what is meant by 'behaviour' here in abstraction from an understanding of the mindedness of the behaving subject. Fodor's cloud chamber analogy brings this out, for in such a case there is clearly no problem in understanding the concept of 'streak' in abstraction from the concept of 'charged molecular particle'. Whether or not the same is true of behaviours such as 'smile', 'laugh', 'wince', 'cry', 'pout', 'stamp', 'grunt' or 'nod' is a different matter.<sup>84</sup>

On the one hand it is clear that behaviours such as smiling cannot be simply anatomically defined. The depiction of a smile as an upwardly curved semicircle is merely a western symbolic convention unrecognised in other parts of the world. Physiologically speaking, certain muscular movements around the mouth may frequently occur in smiles, but the surrounding context of the positioning of the head, general body position, variation in the circumorbital region are also important.<sup>85</sup> And true, smiles may occur in other contexts to indicate, for example, doubt, acceptance, humour, ridicule, superordination,

---

<sup>84</sup> For more on this theme: Mary Midgley, *Beast and Man*, ch. 5.

<sup>85</sup> See Ray Birdwhistell's *Kinesics and Context* ch. 5. On p. 34 we find: 'Even our most preliminary investigation reveals the lateral extension of the corners of the mouth or the upward pull on the upper lips, or any combination of these do not make a recognisable smile. These same activities occur with a snarl or a grimace of pain.'

equality and polite tolerance. Nevertheless, a classic 'warm' smile is by and large one of the essential expressive indices of happiness: not a mere facial movement but something intrinsically expressive.

The inferentialist might admit that many such behaviours (laughs as essentially expressive of mirth, a nod as essentially a communicative gesture and not merely a head movement, and so on) are not intelligible in what they are outside of an intentional, gestural context, but nevertheless suggest that what folk-psychology is really aimed at explaining is the movements and sounds made by people. The trouble with this view, however, is that it is completely wrong. We are scarcely ever interested in the movements that people make - what matters to me are not the means by which you change from one bodily posture to another, but rather the reasons you have for engaging in your intentional actions, and the feelings that you have as displayed in your expressive behaviour. Much of the time I can simply see what it is you feel, or, in viewing the entire stretch of behaviour making up your action, see in the action the intention that defines it as the action it is (as in: putting the kettle on, turning on the light). On occasion the intention may be opaque to me, in which case I may have to ask you what you are doing, which is another way of asking what your intention is. At other times I may not be able to ask you what you are doing, and so be forced to guess; perhaps occasionally I employ a theory.

If this behaviouristic conception of action as mere movement is what is guiding the cognitivist here - and the cognitivist description of action as 'output' would seem to bear this out - it can be neatly compared to the cognitivist's conception of perception as 'input'<sup>86</sup>. Once the proximal content of perception is defined in terms of energy input or sensory stimulation, there is a perceived need for a good deal of 'cognitive processing' to enable the surface stimulations to be turned into what is recognisable as a genuine perception, albeit one located by the cognitivist in the 'inner' world of the mind. Similarly, once action is conceived in terms of output, or behaviouristically, there will automatically appear to be a need to employ elaborate cognitive processes - or, in an intellectualist vein which views practical knowledge as grounded in representational knowledge, a need to employ an elaborate 'theory of mind' - simply in order to recapture the intentionality that has been sapped out of our actions by the theory. What the criteriological conception provides us with is a principled account which avoids both the need for the ordinary subject to constantly apply a vast body of theory (that, intuitively, they do not possess) to their

---

<sup>86</sup> See previous chapter.

experience of others, and also the need for the philosopher to develop such theories as the 'theory-theory' to show how (by analogy or inference) we are to fathom the hearts and minds of others.

## vi. The Desirability of Explanation

The general approach developed above allows us to look again at certain (admittedly miscast) intuitions and arguments concerning psychological explanation that have had little airing since the 1960s.<sup>87</sup> Consider how the issues pan out in Fodor's discussion<sup>88</sup> of David Hamlyn's critique of psychological theories of perception<sup>89</sup>. According to Fodor the intuition behind Hamlyn's critique of certain projects of psychological explanation is a general one; he talks of (p.4):

...those objections to the traditional program [in psychology] that depend upon maintaining that a systematic explanation of *anything* is impossible because, for example, to say that something needs to be explained is to imply that it is somehow untoward, abnormal, and unusual. ... This is a view that has been widely held. It would, for example, seem to be the sort of thing Hamlyn has in mind when he asks [p. 20]:

What of explanations with regard to the veridical perception of the equality of length of a pair of lines seen under normal conditions? There is no question in this case that the lines might be expected to be seen as otherwise than of equal length. There is no question of a deviation from expectation, and if, nevertheless, some question were still asked beginning "Why—," it must be a different sort of question from normal requests for explanation. For explanation is normally called for when the phenomena are regarded, from a certain point of view, as unexpected.

Against this Fodor (rightly, and following Grice) stresses the difference between logical and pragmatic implication:

From the premise "To request an explanation of X is often to imply that X was unexpected" it is concluded that explanations of expected events are *logically* objectionable. But that conclusion follows only if it is also

---

<sup>87</sup> Though for an historically surprising recent addition see Ilham Dilman's *Psychology and Human Behaviour: Is There a Limit to Psychological Explanation*.

<sup>88</sup> Fodor, *Psychological Explanation*, pp. 4–8.

<sup>89</sup> Hamlyn, *The Psychology of Perception*.

assumed that "X requires explanation" *logically* implies "X was unexpected." There are, of course, other possibilities. For example, the implication might be pragmatic (i.e., it might be carried by the speech act of requesting an explanation, rather than by the meaning of "explain").

He then presents (pp. 5–6) an argument that reveals that as a general way of dealing with requests for explanation the principle on offer, when construed as logical rather than pragmatic (or even, for that matter, when construed as pragmatic), leads to cases of 'patent anomaly':

Consider explaining why water beads on smooth surfaces like freshly waxed cars. Presumably the explanation would refer to the reciprocal relation between surface tension, which acts to form the fluid into a sphere, and friction, which tends to inhibit that action: in the "abnormal" case, the effects of the inhibitory force are relatively negligible. Note, however, that if we accept this explanation of the abnormal case, we are ipso facto committed to a corresponding explanation of the behavior of water in the "normal" case, in which water behaves in the way it is expected to behave (that is, in which it does *not* bead). .... In short, the form of the explanation of a phenomenon that we regard as *unusual* implicitly determines the form of the explanation of what we take to be the typical case. We cannot, therefore, hold that the former case is logically capable of explanation and at the same time deny that the latter is too.

Construed in such a general way, the principle that explanation is only called for when we are concerned with what is unusual or unexpected, is false. It confuses a logical sense of 'called for' (i.e. explanation called for simply because it is logically possible) with a pragmatic sense (explanation called for because it is desirable and desired). But whilst Fodor's example shows us that any such general principle is fallacious, it cannot show that the principle as applied to our understanding of the mind is invalid. In what follows I shall argue that Hamlyn's intuition that, say, psychological explanations of our veridical perception of the equality of a pair of lines seen under normal conditions are illegitimate, *is not weakened but rather strengthened* by the failure of his own general principle. So too, Fodor's comparison example of water beading on waxed cars serves not to destroy Hamlyn's intuition, but rather to provide a relevant contrast, a contrast between our understanding of mind and action and our understanding of physical processes.

Let us suppose that desire is a dispositional concept in much the same way that fragility is a dispositional concept, and contrast the two of these with satisfying the predicate 'is yellow' (which is not

dispositional). What I want to suggest is that there is a disparity between the dispositional cases and the non-dispositional case which explains why Hamlyn's intuition applies to psychological cases but not to all non-psychological cases (such as a buttercup's being yellow, or water beading on a car). In short, the argument is that whether or not we are justified in holding that a particular kind of explanation of some expected phenomenon is called for, will depend on whether our expectation is of a piece with our very *understanding* of the phenomenon in question, or whether it is a function of our empirical belief about or knowledge of the phenomenon. I will now explain what I mean.

If I am asked why it is that being fragile makes a glass more likely to break when dropped I should have to return the question without a straightforward answer. I should have to explain what was wrong with the question, and to explain that being likely to break when dropped is at least part of what is meant by 'being fragile'. There is a sense in which an explanation as to why a fragile object did *not* break when dropped will be of a different order than an explanation of why a fragile object is likely to break when dropped. Similarly, sight is the faculty of telling how things are in the environment by using the eyes. It is for this reason that there is no need for explanation as to why someone can perceive the equality of a pair of lines in normal light. In fact whereas it makes perfect sense to ask 'Why was John not able to see what was well lit and right in front of his face?', it barely makes any sense to ask 'Why was John *able* to see what was right in front of his face?'

None of this is to say that there are not a host of important questions to be asked about the mechanics of perception, of *how* it is that we are able to see what we normally do. Hamlyn's own approach does not rule out such possibilities: as he says, when 'There is no question of a deviation from expectation, and ... nevertheless, some question were still asked beginning "Why—," it must be a different sort of question from normal requests for explanation.' The relevant point here is that the type of explanation required is quite different. If an event is expected and the expectation is a function of semantic rather than empirical knowledge then the kind of answer explaining, by the provision of defeating conditions, why it did not happen will be of a quite different order from one explaining how it can happen. At the physiological level similar explanations of both how we are and how on occasion we are not able to act can be given; at the psychological level however the provision of defeating conditions has no comparable complement.

The providing of enabling conditions does not supply new information about the situation in question, but rather reminds us of the semantics of the key terms.

To return to the case of desire. The concept of desire, it was suggested, is a dispositional one. If S desires that O then, all else being equal, they will (it is logically and not merely pragmatically implied) try to bring it about that O. Whilst, then, there may be cause for a psychological explanation as to why John did not procure an ice-cream given his current desire, there is simply no room for a psychological explanation of why, everything else being equal, John *did* act on his desire. The concept of desire has, as it were, done all the work for us already.

If the concept of desire were a concept of an inner item constitutively divorced from behaviour it would, it seems, make sense to ask for an explanation both for how it is and for how it isn't manifest in behaviour. A causal story would perhaps be given which would trace either a causal pathway between the 'inner' desire and 'outer' action, or a causal path from a desire which would stop short of an action. A perfect parity would exist between such cases, and whether not the action were expected given the desire would be irrelevant to the legitimacy of psychological explanation for the action as well as the non-action.

On the other hand, if this take on desire indicates an alienated conception of mind, as the criteriologist maintains, and if desire is constitutively related to action and is not a self-contained inner state, the desirability of psychological explanation for the normal case is instantly called into question. My expectation that you will act on your desire is not one based on experience or derived from a theory, but rather grounded in my understanding of what it is to have a desire in the first place. It is because of the imminence of the mind in behaviour, and not because of a general principle covering all cases of explanation, that Hamlyn was right to reject *psychological* explanations of normal behaviour. The present case illustrates perfectly the Wittgensteinian theme taken up in Part 1: The belief that explanation in philosophy is fundamentally misguided need not be thought simply a matter of taste, but rather an insight into the fact that philosophical explanations are typically consequent upon (perhaps implicit) conceptions of the subject matter in hand which portray it in a fundamentally alienated or disengaged light.

## **vii. Summing Up**



Let us consider the argument so far. Section 2 began with an explication of Wittgenstein's criteriological conception of agency – of mind in action – which stressed both the constitutive rôle that expression and action play in the individuation of mental contents, and the happy epistemological consequence of this rôle: that mental contents can be ascribed to others on an observational basis without the need for inference to a hidden inner realm of mental entities. This explication was developed in the context of a critique of the mentalist's analogical view – that we understand what it is for others to have mental contents on the basis of our understanding of what it is for ourselves to have mental contents.

Fodor presents Wittgenstein's criteriological conception in a not-unfair manner, but suggests that there exist various weaknesses which can after all be better dealt with by a form of mentalism which uses theory rather than analogy to bridge the link between the supposed gap between the inner mental realm and the outer behavioural realm. These weaknesses however were argued to pertain more to Fodor's exposition than to the criteriological conception itself. Fodor's 'inference to the best explanation' theory was then itself subjected to criticism. First of all it was noted that mental phenomena do not always stand in a fundamentally explanatory relation to behaviour. Second it was questioned whether the inferential conception truly possessed the resources to show how mental phenomena can stand in such an explanatory relation.

Third it was argued that the degree of logical disconnection between behavioural and psychological concepts required by the view that our appreciation of others is grounded in inference, theory, and explanation is simply not present. On the one hand our everyday *expectations* concerning what others will do do not have to be modelled as quasi-scientific *predictions*. And an examination of such expectations shows that our understanding of the minds of others is simply of a piece with (rather than provides the grounds for) our appreciation of their character – of their likely actions. (Hence psychological concepts *are* internally related to behavioural concepts.) On the other hand our everyday actions and expressions, at least, those with which we (qua 'folk-psychologists') are typically preoccupied, are not to be construed as mere movements, but rather as intrinsically mental.

Finally the implications of a criteriological conception of agency for psychological (rather than folk-psychological) explanations of action and perception were sketched. Here the general theme of the first

two Parts of the dissertation was reintroduced: it is only on the assumption of an alienated conception of the mind – that is, of a conception of mind as not imminent within, but rather hidden behind, action – that epistemological theories concerning *how* we exercise our psychological faculties gain any kind of ascendancy. Wittgenstein's and Ryle's argument with epistemological theories of *how* we think, perceive and act need not be seen – as Fodor (perhaps correctly) sees Hamlyn's argument – as a product of outdated positivistic or reductive conceptions of meaning, but rather as a reaction against reductive mentalist conceptions of mind which reduce the logical form of talk about minds to that of talk about bodies.

### 3. On Expressing Ourselves

So far it is only the metaphysical relation of mind to action and the correlative epistemology of our knowledge of other minds that has been discussed. Equally important are a set of first- rather than third-personal questions concerning the logical and epistemological status of self-ascription and self-knowledge.

Delimiting a distinctly cognitivist account of psychological self-ascription is none too easy, and this largely because the whole of the debate surrounding the topic is couched in terms which from the perspective being developed here invite diagnosis rather than frame the entire spectrum of possibilities. So it is often thought that our capacity to self-ascribe a pain, say, is made possible by first person *access* that we enjoy to the sensations and thoughts that 'occur in us'<sup>90</sup>. The general faculty underlying this access is known as 'introspection', which is sometimes explained further as a form of 'inner sense' or

---

<sup>90</sup> Sydney Shoemaker's entry on 'introspection' in *A Companion to the Philosophy of Mind* has it that (p. 395) 'In its broadest sense it refers to the non-inferential access each person has to a variety of current mental states and events - sensations, feelings, thoughts, etc. - occurring in that person'. The five pages of discussion that follow take this definition as providing the parameters of the debate; the 'access' metaphor is not interrogated (although its metaphorical character at least casts doubt on the possibility of using the term in an explanatory way); the correlative metaphysics to the 'access' epistemology - the idea that feelings and thoughts occur *in* people - is simply taken for granted, as is the very idea that our knowledge of what we think and feel is typically provided by introspection.

'self-consciousness'. In cognitive psychology the idea is frequently propounded that our capacity to self-ascribe is made possible by the activity of a 'self-monitoring mechanism'.<sup>91</sup> In one form or another contemporary debates frequently take it for granted that I can be said to be 'conscious' of my own 'mental states' (my beliefs and hopes etc.).

It is for this reason that what follows focuses on the undeniable reality of psychological *self-ascription*<sup>92</sup> rather than any putatively general *mental acts* or faculties which underpin this capacity. The conception of mind that is styled cognitivist is one which supposes that our capacity to self-ascribe is dependent upon our having 'epistemic access' to our own thoughts and feelings and sensations. In the most extreme cases this 'access' is modelled in perceptual or quasi-perceptual terms - i.e. as a faculty of inner sense. But in the most general case the idea is that when I say what I think or feel, my utterance normally has the character of a *report*: I let on to you what, in introspection, I find to be the case with me.

To consider some examples: David Armstrong provides a cognitivist treatment of self-ascription in his *A Materialist Theory of Mind* (p. 84). According to Armstrong, introspection is a process whereby the brain scans itself in a way that somehow provides for awareness of some of its present states. Whether or not this is adequate as a theory of introspection is not the issue I wish to raise; the relevant point is just that Armstrong assumes our capacity to say what is on our minds and what we feel to be underwritten by introspection, however physically realised. Similarly Patricia Churchland (*Neurophilosophy*, pp. 305 ff.) takes it upon herself to dispute the suggestion that 'the attribution to *ourselves* [of folk-psychological states] seems directly observational, without any mediating framework of categories and concepts ... surely one has direct and unmediated awareness of one's own mental states.' Again, Churchland's alternative suggestion that our self-ascriptions are instead mediated by theory is not the relevant issue here. What is more pertinent is that both the 'direct' and the 'indirect' alternatives are framed as forms of 'introspection' which deliver accurate or inaccurate 'inner perceptions'. Both authors presuppose that self-ascription is grounded in an epistemic faculty of inner sense and fail to consider non-cognitivist alternatives.

---

<sup>91</sup> cf Weiskrantz, *Blindsight*; Frith, *The Cognitive Neuropsychology of Schizophrenia*.

<sup>92</sup> By 'self-ascription' I mean *whatever* it is we do when we say 'I feel X', 'I think Y'; I do not mean to suggest that such utterances express judgements or offer descriptions.

### i. Problems with Cognitivism

A question that immediately arises for the cognitivist is whether the introspective model can really do justice to the phenomenological facts. For although, whilst enjoying a theoretical and speculative (and *disengaged*) moment, the idea that we discern the contents of our own minds by a form of *looking within* is relatively appealing, this activity is scarcely notable for its occurrence in those everyday situations when we simply say what we think. That is, it is scarcely notable when we 'speak our minds', avow an intention, let our feelings be known, give voice to an opinion, say how we feel, and so on. The immediacy with which we apprehend and advertise our thoughts and feelings is not captured, in any straightforward manner, by the introspective theory. We just say what we think *without* first having to find out; or at least, so it appears.

This objection could be obviated by rendering the operation of the introspection or of the self-monitoring mechanism unconscious, but this manoeuvre makes it incumbent on the cognitivist to provide *evidence* that such a mechanism exists. Aside from the fact that we can and do self-ascribe, it is hard to see what would count as such evidence. If the cognitivist's theory were the only one in town, it would perhaps have to be accepted as it is. But there are other accounts of self-ascription available (described in subsection 2 below), and further compelling arguments against variants of the cognitivist's theory.

The two principle objections to cognitivism about self-ascription focus on the cognitivist's claim that our self-ascriptions are *judgements*. The first argument will be that cognitivism cannot accommodate an essential aspect of our subjectivity, namely, the *first-person authority* we possess concerning our own feelings and thoughts; the second will be that cognitivism unwittingly embarks us on a regress of judgements about our own minds.

In the previous section (2) it was argued that desire and action were not just empirically found together in regular causal conjunction, but are rather mutually constitutive phenomena. It is not intelligible, so the argument went, that we may desire something yet, in the absence of defeating conditions, not act upon it. Now a similar lack of intelligibility seems to characterise the suggestion that

we characteristically err in appreciating what we believe, or err in our self-ascription of a toothache or an intention. To be sure, to the extent that the world itself enters into our psychological 'state' - as in the case of knowledge - we can be wrong. I may self-ascribe knowledge when, given that things are not as I suppose, I ought to have self-ascribed belief. Nevertheless the self-ascription of belief (or pain or intention) does not appear to be fallible in the same way. If I am saying what I think or feel or how things seem to me then I cannot sensibly be thought to err - this logical possibility is not open to me.

The mentalist's account of self-ascription, however, only makes it *more or less likely* that I should not err in self-ascribing, and leaves it as a logical possibility that I might only think that I have a headache when in fact I don't. The Cartesian supposes that I do not tend to err because I enjoy a 'clear and distinct perception' of my own sensations; the cognitivist supposes that the 'monitoring mechanism' which underpins my 'self-awareness' is generally working well. Neither of these options, however, seems to provide for the kind of authority we possess in making self-ascriptions: they both repudiate our right *qua* subjects to be treated as the authority on what we think and feel, and turn the constitutive inalienability from us of our own mental contents into a contingent non-alienability. In so doing they not only force a conceptual wedge between a subject and his own mind, generating a kind of self-alienation, and thereby sanctioning the logical possibility that I may mistake a toothache for a whim, or a belief in democracy for a desire for an avocado, but also leave us with a dearth of criteria for the ascription of the (mistakenly) self-ascribed contents in the first place.

To elaborate these two points: First, the cognitivist model makes it logically possible that I be mistaken about what I think or feel, although when we are considering everyday<sup>93</sup> thoughts it not only seems highly unlikely but rather completely unintelligible that I should be mistaken about what I think. In the normal run of things I am to be treated as the authority over my own mind, and if someone else questions whether I really think what I self-ascribe I may rightly be indignant and offended at this seeming lack of respect of my status *qua* person. Secondly, for some of the more complex 'propositional attitudes' the principle criterion for attributing them to a subject seems to be that subject's willingness to self-ascribe the attitudes in question. For such attitudes only linguistic behaviour (and not bodily actions

---

<sup>93</sup> That is, not unconscious affects or unconscious phantasies.

or bodily expressions) will have the requisite conceptual complexity to enable the attribution to be sustained with warrant, and the principle such behaviour is the self-ascription of the attitude.<sup>94</sup>

In pursuing their introspective analysis of self-ascription, the cognitivist ends up both alienating us from ourselves, and also failing to take us seriously as people. In denying us the authority we have concerning our own thoughts and feelings which is granted us by grammar<sup>95</sup>, the cognitivist estranges us from not only the world and other people (as suggested above), but also from our own selves, i.e. from ourselves.

The above represents one arm of the logical support for what is intuitively the phenomenological failure of the cognitivist position to account for self-ascription. The second arm makes reference not to our authority as subjects, but to the grammatical character of the self-ascription.<sup>96</sup> If the cognitivist is right, then when we self-ascribe a propositional attitude, we are offering a report of what we find inside us, a report which could perhaps be mistaken. That is to say, on the cognitivist's account, our self-ascription can be seen as the expression of a judgement - for it is judgements that can be correct or incorrect. The judgement tells someone else what we think or believe is going on in our minds. In the normal run of things, so long as our self-monitoring equipment is in order, our judgement will be correct, and we will have the propositional attitudes that we say we do. It is logically conceivable, however, that the monitoring system breaks down (it is just this logical space which is occupied by the theory of self-monitoring mechanisms), and in such cases our belief that we have attitude X within us will be mistaken.

That this is a questionable conception of mind can be seen from considering the self-ascription of epistemic attitudes. If what the cognitivist asserts is true, then when I say that I believe that there is a cat on the lawn, what I am really doing is expressing my belief that I believe there is a cat on the lawn. My

---

<sup>94</sup> Many complex propositional attitudes are perhaps ascribable on the basis of what else is said: my belief that X will be the next president may be inferable from what else I say; other criteria are therefore available. This is not however always the case: various emotions or feelings, for example, are characterised by terms used in a 'secondary sense' (see ch. 7), where the criterion for the feeling being correctly described as 'as of Y' is simply the subject's assent that this is the correct description.

<sup>95</sup> i.e. in virtue of what it *means* to be a subject.

<sup>96</sup> The following is inspired by Rockney Jacobsen's paper *Wittgenstein on Self-Knowledge and Self-Expression*. See also Dorit Barron & Douglas Long's *Avowals and First-Person Privilege*.

self-ascription of a belief is itself a fallible judgement I make about myself - I suppose that I believe that there is a cat on the lawn - but I could (logically) be wrong. But now what should we say about the subject's epistemic relation to the content of this second-order judgement? If the self-monitoring theory is on the right track, then it would seem that I cannot simply and directly express this belief about my belief. I may after all be wrong in thinking that I believe that I believe that p. Perhaps I do not really believe that I believe ... In this case, when I say that I believe that p, the logical form of this proposition should be spelt out not as 'I believe (or know) that I believe that p', but rather as 'I believe (or know) that I believe (or know) that p'. But the question still arises: Do we or do we not have an epistemic relation to the third-order belief? If we do then it is presumably underpinned by the cognitivist's self-monitoring mechanism, is something found out about by introspection. My self-ascription of a belief, then, represents an assertion about a state I find in me, which state is a belief that I have in me another such state, the content of which is that I have in me another such state, and so on *ad infinitum*.

The regress is clearly vicious and it is tempting therefore to jump off at some convenient point. A natural place for the cognitivist to disembark is at the level of the second-order beliefs: My self-ascription of a belief (in 'I believe there is a cat on the lawn') expresses not a belief that I believe that I believe there is a cat on the lawn, but rather just my belief that I believe there is a cat on the lawn. It lets be known my judgement, delivered by introspection, that I have such and such a belief inside me.

This is all very well, but it is hard to see how the cognitivist can justify their choice of a stopping place. On the one hand the choice of the second-order level seems arbitrary, and merely prescribed by the theory. But on the other hand, the fact that it is allowed that a self-ascription can directly express, rather than report on, a belief, albeit one of the second-order, seemingly undermines the motivation behind the cognitivist's original claim, that self-ascriptions cannot simply express the attitudes that they ascribe, but rather give voice to an epistemic attitude toward the attitude in question. For if a self-ascription can simply express rather than report on a second-order belief, why can't it do so for a first-order belief? Why shouldn't a self-ascription of a belief be seen simply as an expression of that very belief itself?

These objections - from phenomenology, authority and grammar - have it that cognitivism provides us with an alienated *a priori* psychology, one which distances us from ourselves in a way which (if the regress argument is correct) sets us not only at one, but rather at an infinite, remove from that - our own

minds - to which we have traditionally been thought closest. In supposing that our self-ascriptions express beliefs about what lies within us, the cognitivist opens up the requirement for a theory of how we come by such beliefs, a theory which is provided in the form of posits of a faculty of inner sense, introspection, or of a self-monitoring mechanism. Once again, the psychological 'how?' questions ('how do we know what we believe/feel/intend?') can be seen to arise posterior and not prior to the provision of an alienated account of mindedness.

The suggestion that the modern conception of the self is fundamentally alienated is hardly a new one.<sup>97</sup> But what is striking from the point of view I have adopted here is the way in which cognitivism consistently compounds this alienation, resulting in a kind of double disengagement, or an alienation squared. First the mind is alienated from the lived body: the desires and intentions of the self are not seen as imminent within action but as disengaged occupants of an inner mind space. Perception is theorised on the model of sensation: the psychologically essential elements of perception are held to be dissociable from the environment - in short, perception is modelled on sensation. But then sensation is itself modelled on perception - we are held to be in an epistemic quasi-perceptual relation to our own sensations. The analysis generalises for all of our so-called 'self-knowledge' or 'introspection': our capacity to let on what we feel is modelled as a capacity to describe what we inwardly perceive to be happening in our inner realm. Along the way the true self has retreated first behind the body into an inner mind space, and then second behind this mind which is represented as a realm of objects into which the true self can peer.

This in itself is not however the complete diagnosis, for mention must be made again of that secret smuggling back in of conditions of intelligibility within a context that explicitly denies them - the fallacy earlier referred to as 'Newton's error'. Volitional theories presuppose inner *acts* in their theory of action; representational theories of vision presuppose something like inner pictures which require to be *seen*; representational theories of content presuppose the *normativity* that these inner representations are supposed to explain; rule-based theories of rationality presuppose the capacity to interpret the rules in a

---

<sup>97</sup> Cf Martin Heidegger, *Being and Time*, Richard Rorty, *Philosophy and the Mirror of Nature*, Charles Taylor, *Sources of the Self*.

Marx's early writings on alienation have been important for the examination of this theme in the social sciences. cf Richard Schacht, *Alienation*, and Alasdair MacIntyre's *Against the Self-Image of the Age*.



*rational* manner, know-how is consistently explained in terms of knowledge although the ascription conditions for the knowledge always presuppose a vast amount of *know-how*; inferential conceptions of our understanding of mind presuppose the concepts of the *mental contents* to which inferences are made; finally, cognitive theories of self-ascription which explain expression in terms of assertion presuppose all along the capacity to *express*.

There is in a sense little need to develop an alternative conception of self-ascription to replace that offered by the cognitivist. This is because the cognitivist conception itself, in a way that is becoming familiar, must itself (illicitly) presuppose that we are able to simply *avow* our own mental states themselves and not always be restricted to occupying an alienated reflective position with respect to the contents of our own minds. The alternative 'avowalist' conception is, in terms of its theoretical significance, ultimately a purely *negative* doctrine. It does not aim to explain *how* we can authoritatively and securely self-ascribe, for this question itself seems to presuppose that such self-ascription represents some form of epistemic achievement, and this presupposition is of a piece with an alienated conception of mind. There is however a need for an account of the logical grammar of first-person folk-psychological ascriptions, and in what follows such an account will be provided along with suggestions as to how to avoid misleading implications generated by analogies commonly employed to explicate the 'direct' (non-mediated by judgement) character of the avowal.

## ii. Avowalism

It is hard to imagine opining that a mouse squeals, a pig grunts or a baby cries in pain because they have decided that that is what they feel and wish to convey how it is with them to some other organism. Such noises are rather direct *expressions* of the pain, and are not to be modelled as *assertions* that it is pain that we have to do with here. It is from such observations that the philosophical position known as *expressivism* takes its lead. Although such animal utterances are not linguistic, self-ascriptions of sensations using language can be thought of as akin to these noises: they are the linguistic surrogates of such animal expressions. Something like this position seems to have been in Wittgenstein's mind when in *PI* §244 he considers how it is that people learn the meaning of 'pain':

Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behaviour.

The non-epistemic attractions of this model, in accounting for the immediacy and immunity to error of psychological self-ascriptions, should be clear. For grunts and squeals are not judgements and are not the product of introspection. Nevertheless there are clear dangers that result from an over-assimilation of the articulate self-ascription of a propositional attitude to the model of animal expression. For whilst it seems reasonable to suggest that avowals are not typically to be thought of as correct or incorrect, given that they are not always employed in the expression of second-order judgements, it seems far-fetched to urge that, like inarticulate animal expressions, they cannot be true or false. Furthermore, whilst animal expressions are typically reflex-ive, avowals can be both reflex-ive and reflect-ive, both spontaneous or automatic and deliberative or consciously entered into. Finally, whilst it has been argued above that there are multiple defeasible behavioural *criteria* for ascribing mental states etc. to subjects, avowalism presents us with a conception of self-ascription which maintains that self-ascriptions are *not* subject to criteria of correctness. In all three cases the expressivist treatment of avowal seems to purchase the inalienability and authority of self-ascriptions at the price of their articulate, propositional, truth-bearing, normative status.

All three objections are mistaken; in what follows I shall reply to them in turn.

First, since avowals (unlike animal noises and one-word exclamations) are fully syntactically and semantically fit ascriptions, they are perfectly suited to the task of truth-bearing. 'I am in pain' is true iff I am in pain whether or not the sentence constitutes a report or an expression. There is no reason to think that the only expressions which can be described as true or false are expressions of belief, and it is simply a fact that articulate expressions of sensations or emotions or intentions such as 'I intend to go to bed', 'I have toothache' – just as much as expressions of beliefs – will be true or false depending on whether the subject has the intention or the pain in question.

The account given does however rule out the possibility of *error* in self-ascription. And it does this from a grammatical perspective, which is to say that it characterises the language-game of avowal as one

without a logical space for talk of correctness and incorrectness. This however seems to conflict with the possibility of avowals being true and false – for are not truth and correctness one and the same? The simple answer is that truth and correctness are *not* one and the same. It is judgements, the beliefs they express, and the people who express them that are aptly described as being correct or incorrect. So, to be sure, when what is avowed is a belief, the avowal can be thought of as correct or incorrect depending on whether the belief avowed is correct or incorrect. When what is avowed is not a belief, however, talk of correctness or incorrectness is simply out of order.

Even here there is an important need to qualify this claim which further demonstrates the logical distance between the notions of truth and correctness. An avowal of a belief is correct or incorrect depending on whether *the world* is as the judgement depicts it to be. In the sense, then, in which a self-ascription of a *sensation* may be true or false depending on whether the subject truly has the sensation in question, an avowal of a belief may be true *even though the belief be false*. That is to say, the avowal will (in this sense) be true iff I have the belief in question, but will only be correct if things are as I believe<sup>98</sup>. The only times truth and correctness will run together in the way in which the objection requires will be when my self-ascription of a mental state amounts to an avowal of a belief that I have the mental state in question. Whilst this may occasionally happen (in a psychoanalytic setting, for example) it is clearly unusual, depending as it does on taking a third-person perspective toward ourselves which, if it were the norm, would be indicative of a highly alienated consciousness.

The second objection noted that linguistically structured avowals – unlike animal grunts – are capable of being issued deliberately. I may for example make a point of showing you my gratitude; whilst sometimes my gratitude is spontaneously expressed with a phrase such as ‘I am so grateful’, at other times I, after reflection, make the self-ascription with a deliberate intention. (Perhaps I know that my gratitude would mean a lot to you.) Even here, however, the presence of the intention need not be taken as indicative that the self-ascription expresses a belief or advertises a judgement that I have the feeling in question. This is obviously true when we consider non-verbal gestures: I may deliberately express my gratitude with a hug, but this is not to say that my hug expresses a judgement that I am

---

<sup>98</sup> C.f. *PI* §290.

grateful rather than my gratitude itself.<sup>99</sup> When I show you how I feel, even after having decided so to do, I let my feelings be known by expressing *them*, not by expressing *judgements about* them.

The third-objection pointed out an apparent discrepancy between the criteriological conception of mental states (as criterially related to behaviour) and the avowalist's conception of self-ascription (as not subject to criteria of correctness). The discrepancy however *is* merely apparent. My avowal of a pain is not correct or incorrect because it is not a judgement: the *self-ascription* is not correct or incorrect. But of course the *concepts* employed in making the self-ascription are governed by criteria, criteria which I must master if I am to use the word 'pain' etc. correctly. If I express a wild feeling of joy my self-ascription cannot be correct or incorrect because feelings of joy cannot be correct or incorrect and my self-ascription expresses such a feeling and not a belief that I have such a feeling. Nevertheless I may not have mastery of the concept of joy, in which case my self-ascription is unlikely to constitute an expression of joy. (I may think that 'joy' means misery, and so self-ascribe joy when I am trying to express my misery.)

Attempts to spell out the expressive (rather than assertoric) character of avowals have typically been criticised for denuding the first-person uses of psychological terms of their essential normative and semantic properties. The account above does not however attempt to model avowals on the grammar of anything other than themselves; what has rather been of fundamental import is avoiding the regress invited by cognitivism consequent upon the supposition that avowals only ever really express judgements or beliefs.

The avowalist account is in fact rather helpful in elucidating the essential complementarity of first and third person with respect to folk-psychological ascriptions. On the one hand our epistemic relations to others are rendered explicable by an approach which views self-ascriptions as expressive rather than assertoric. Just as we can directly relate to someone's pain as expressed in their grimace without depending on an inference to an inner world, so we can directly concern ourselves with the *content* of a self-ascription without first wondering if the self-ascription is correct. Just as your laugh can function as

---

<sup>99</sup> C.f. Bar-On & Long *op cit.* p. 330.

a criterion for me for judging that you are amused, so too can your self-ascription. It is because of the kinship of avowals to expressions in general (and not to assertions) that this is possible.

On the other hand there are important respects in which it is the *differences* between first and third person that reveal the essential complementarity in question. For example, according to Wittgenstein: 'Not until [a psychological term] finds its particular use in the first person does it acquire the meaning of mental activity' (RPP II §230). If this is true then the conceptual richness of third-person folk-psychological ascriptions is dependent on their being able to find a first-person employment. To give another example: it is not unreasonable, if we are to credit someone with an understanding of the terms that they use, to request justification from them for what they say. This will hardly be possible however in first-person employments of the terms: my avowal 'I am happy', being an avowal and not an assertion, is not based on grounds of any sort. It is however possible in third-person ascriptions, where my understanding can be manifest in justifications such as 'I saw her smile'. What semantically underwrites my use of psychological terms in self-ascriptions, then, is (arguably) my capacity to employ them in other circumstances as well.

A detailed description of the logical grammar of self-ascriptions and their relation to other-ascriptions must wait another day. This however is not too troublesome, for what has been demonstrated is that cognitivism not only employs what has been urged is an alienated as opposed to immediate and lived conception of the subject (an observer rather than participant model of self-ascription), but that, if it does not wish to embark on a never-ending regress, cognitivism must after all presuppose that we *can* simply – without the mediation of judgement – avow our own mental contents. Once again it is 'Newton's error' that best captures the logical form of the theoretical manoeuvre in question. An attempt is made to give a substantive answer to a question such as 'How is self-ascription possible?', but the question presupposes an alienated conception of the topic in hand. The aim was to psychologically explain how we make our most basic self-ascriptions – our avowals – and it was suggested that such self-ascriptions are assertions concerning what we find within us. The capacity to make an assertion, however, presupposes that one can – after all – simply avow what one believes to be the case. Once again what was supposed to be being theorised has tacitly and illicitly turns up within the explanation. Or to put

it differently, the spontaneous, praxical and expressive subject tacitly and illicitly turns up *within* the deliberative, reflective and disengaged subject as theorised by cognitivism.

### iii. Summing Up

Part 2 of this dissertation has been concerned to argue against the cognitivist's account of mind, action, perception and thought. Chapter 2 argued that the cognitivist asks a variety of questions concerning how we perceive, recognise, think, draw inferences etc., but that such questions only appear intelligible when approached from a perspective which construes the subject as alienated from the world. Chapter 3 extended this argument to epistemological issues: the cognitivist asks how it is that we know the minds of others, and how it is that we know our own minds. Both questions however presuppose epistemic gulfs which are unreal. The view that our understanding of others involves inference, prediction and quasi-theoretical explanation presupposes a conception of the mind as an inner realm constitutively divorced from merely 'outer' behaviour. The view that our capacity to self-ascribe manifests in the offering of judgements about our own mental states is also alienated, neglecting the natural possibility that we can actually *occupy* or *speak from* our desires and hopes rather than merely offer second-order opinions about them.

Part 3 will concern itself critically with cognitive theories of schizophrenia. Here the arguments of Part 2 are put to work, the principle criticism being that the cognitive theories tend to theorise schizophrenic disorder as a disruption in causal processes or epistemic faculties which are themselves only the misguided posits of cognitivist theory.

## **Part 3**

# **Understanding Psychosis**

## Ch. 4 Cognitive Theories of Schizophrenic Fragmentation

In the following pages, we shall be concerned specifically with people who experience themselves as automata, as robots, as bits of machinery, or even as animals. Such persons are rightly regarded as crazy. Yet why do we not regard a theory that seeks to transmute persons into automata or animals as equally crazy?<sup>100</sup>

### 1. Introduction: Mind and Mechanism in Schizophrenia and Cognitivism

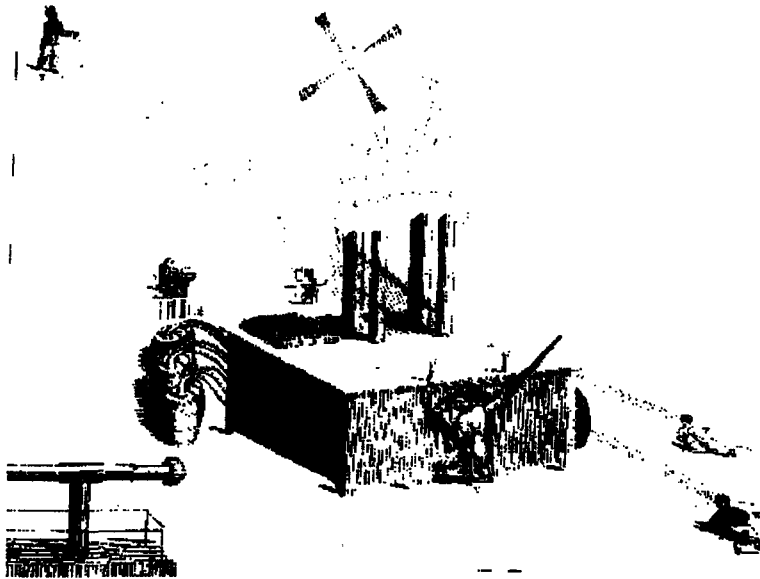


Fig 4.0

#### i. The Influencing Machine in Schizophrenia

Part 3 of this dissertation makes the transition from purely philosophical psychology to questions of psychopathology. The pathology in question is *schizophrenia*, many of the distinctive features of which are illustrated by the well-known delusion of the *influencing machine*. Subjects with this delusion suffer the aberrant belief that they are under the influence of a machine – often one manipulated by malicious agents – that tries to control their thoughts, feelings and actions. The delusion undoubtedly has a basis in

<sup>100</sup> Laing, *The Divided Self*, p.23.



experience, the experiences in question being recognisable to a psychiatrist as Schneiderian first rank symptoms<sup>101</sup>. Such symptoms include thought insertion, broadcast and withdrawal (The schizophrenic feels that thoughts are being put into or taken out of their mind), and passivity experiences (movements, emotions and sensations experienced as caused by an external agency).<sup>102</sup> In what was probably the first clinical description of a patient with (what today would be recognised as) schizophrenia, John Haslam described the 'air loom' by which his patient James Tilly Matthews believed he was being tormented.<sup>103</sup>

In Matthews' case, the machine was experienced as quite different in structure and position from his person. But other common delusions, as the quote from R. D. Laing prefacing the chapter indicates, include the belief that the subject's own body contains a machine – that if, for example, they were to cut open their skin (which may even be attempted), wires and pistons would be exposed. An intermediary case would be that of Miss Natalija A., the subject of Victor Tausk's famous paper *On the Origin of the 'Influencing Machine' in Schizophrenia*.<sup>104</sup> Her influencing machine had the shape of the human body, and everything that happened to this machine also happened to her. In all such influencing machine delusions, as well as in many other such typically schizophrenic delusions, we find experiences of depersonalisation, of mechanisation, of a lack of agency and free will, and of a subjugation to causal law.

Whilst delusions are usually (but inadvisably) defined as having culturally atypical content, the mechanisms by which influencing machines have been believed to operate have certainly reflected the technological developments and obsessions of the age. Natalija A's machine, for example, was powered by electricity, whilst Matthews' was powered by air.<sup>105</sup> Others may, for example, use radio waves or magnetism.

---

<sup>101</sup> Which play a central rôle in the diagnosis of schizophrenia.

<sup>102</sup> Other schizophrenic symptoms of diagnostic importance include: hearing voices, thought disorder, delusions, emotional incongruity, diminished action, speech and feeling, and (less commonly) catatonic and stereotyped behaviour.

<sup>103</sup> *Illustrations of Madness*, written in 1810. It is Matthews' own fine drawing of his Air Loom that is reproduced in the text. Matthews appears in the top left corner; the other people are operators of the machine – 'a gang of villains profoundly skilled in Pneumatic Chemistry'.

<sup>104</sup> 1919. Referred to by Sass, *Madness and Modernism*, pp. 217–8.

<sup>105</sup> Op. cit. p. 19.

## ii. The Influencing Machine in Psychology

The same technological inspiration can also be said to hold for another set of 'influencing machines', the most recent of which is the computer, but which in the past have included hydraulical and other mechanical devices<sup>106</sup>, that over the decades have influenced psychologists' understanding of the nature of the mind. This parallel, between the psychologist's framework for understanding the mind and the schizophrenic's experience of mindedness, has not gone unnoticed in the literature. It is a prominent theme, for example, in both Laing's *The Divided Self* and Louis Sass's *Madness and Modernism*. A brief examination of these phenomenological texts will help situate the analytical critique to follow.

Sass's work in particular aims to promote psychological understanding of schizophrenia by means of an analysis of modernity. Modernism, as Sass characterises it, provides us with a conception of i) the world in relation to the subject as subjectivised and dependent, of ii) the subject in relation to the world as desocialised, disengaged and autonomous, and iii) of the subject in relation to itself as objectivised.<sup>107</sup> And schizophrenia, in Sass's scheme, is as it were the embodiment of this ontological picture in the life of the subject. For such people the world can i) be experienced as being *dependent on their continued representation* of it (idealism), other people ii) may be experienced as *unreal or as mere automata* (solipsism), and iii) their own minds come to be experienced as *objects* - i.e. as *other* (introspective alienation). In short, the delusions of the schizophrenic subject are interpreted by Sass not as unusual *ontical* (i.e. empirical) beliefs about this or that particular thing, but as expressions – albeit in an ontical mode – of a fundamental *ontological* transformation, a pervasive structural change in the way the world is understood, related to, and inhabited.

The parallels between schizophrenia and modernism are not for the moment of direct concern, although the theory of schizophrenia sketched in the following chapters attempts to investigate just what

---

<sup>106</sup> Turing's ticker tape machine for example.

<sup>107</sup> The last two of these components of a 'modernist' conception of the self should be recognisable in the sketch of cognitivism in the

the intuitive characterisation of psychosis as *a state of 'frank alienation'*<sup>108</sup> amounts to. Of greater pertinence is the manner in which cognitive theories of psychosis embody the alienated conception of the self, and the influence that such a conception or preconception about the nature of mind has for understanding the nature of mental dis-integration. In *The Divided Self* Laing writes (p.19):

The most serious objection to the technical vocabulary currently used to describe psychiatric patients is that it consists of words which split man up verbally in a way which is analogous to the existential splits we have to describe here. But we cannot give an adequate account of the existential splits unless we can begin from the concept of a unitary whole, and no such concept exists, nor can any such concept be expressed, within the current language system of psychiatry or psycho-analysis<sup>109</sup> ...

... or cognitive psychology. At least, this is what shall be argued.

In the passage immediately preceding that which heads this chapter, Laing wrote:

It seems extraordinary that whereas the physical and biological sciences of it-processes have generally won the day against tendencies to personalize the world of things or to read human intentions into the animal world, an authentic science of persons has hardly got started by reason of the inveterate tendency to depersonalize or reify persons.

I have already traced such depersonalising, reifying and mechanistic tendencies within (for example) cognitivism's conception of action as output and perception as input. The question remains however whether this reductive tendency is the entire source of the failures of traditional mentalistic attempts to

---

<sup>108</sup> This is how psychotic disorders were known to the 'alienist' precursors of today's psychiatrists. (See Roy Porter's *Mind-Forg'd Manacles*.)

<sup>109</sup> The text continues: 'The words of the current technical vocabulary either refer to man in isolation from the other and the world, that is, as an entity not *essentially* 'in relation to' the other and in a world, or they refer to falsely substantialized aspects of this isolated entity. ... Instead of the original bond of *I* and *You*, we take a single man in isolation and conceptualize his various aspects into 'the ego', 'the superego', and 'the id'. ... This difficulty faces not only classical Freudian metapsychology but equally any theory that begins with man or a part of man abstracted from his relation with the other in his world.' I should (optimistically) like this thesis to be read as an analytical transposition of this aspect of Laing's critique of psychoanalysis into a critique of cognitive psychology.

understand both rational mindedness and schizophrenia. In what follows (as in the two sections above) I shall argue that an adequate explanation of the failures of cognitivist theorising concerning schizophrenia requires that attention be paid not only to the way in which cognitivism tends to provide *reductive* accounts of mind, but also to the ways it tends to *hyperbolic* accounts of the same. That is, not only does cognitivism *under-attribute agency and subjectivity* to the person, but in a mutually dependent way also *over-attributes* agency and subjectivity. In Laing's terminology, cognitivism not only depersonalises persons, but also reads human intentions into the human animal at impossible junctures, personalising the sub-components of our cognitive systems.

This may appear paradoxical, but the appearance of paradox can be removed once the mechanism and reification of cognitivism are understood not simply as a function of cognitivism's naturalistic *reductionism*, but rather as a product of the *alienated* conception of mind it embodies. Take representationalism as an example. On the one hand the representationalist tends to a reductive account of perception<sup>110</sup>, viewing it in terms of mechanical 'it-' processes consequent upon 'input' to the perceptual organs. In this way what is essential to perception is lost. But on the other hand the inner processes are (if the arguments of Part 2 were along the right lines) required for the translation of surface stimulations into inner representations, in other words, into *perceptibilia* for an alienated and internalised subject. Similarly with representationalist accounts of meaning: on the one hand an overly static conception of meaning or understanding is provided which abstracts it away from the activity and life of the subject; on the other it deploys in its theory a concept of an item (a representation) *which itself requires to be understood* in a certain way<sup>111</sup>. Wittgenstein's rule-following considerations make a similar point, as for example does Daniel Dennett's attempt to expunge the 'Cartesian theatre' out of theories of consciousness.<sup>112</sup>

The influence of the machine in cognitive theory, then, is a negative one not simply because it encourages a *mechanistic* conception of the human being, but because it employs an alienated conception of the subject in which the subject has retreated behind the body and even behind the mind. This subject

---

<sup>110</sup> See chapter 2, subsection 2.

<sup>111</sup> See chapter 2, subsection 3.

<sup>112</sup> In *Consciousness Explained*.

then requires to be re-linked with the world. Cognitive theories, supposing that the predicament of such a subject is the predicament of us all, provide what are promoted as explanations of thought and perception that are geared up to relinking such a subject with the world. And in the process concepts are employed – representation, propositional or representational knowledge, inner volitional acts – which presuppose – and so are in any case incapable of explaining – our fundamental psychological capacities and contact with the world (praxis, practical knowledge, intentionality).

### iii. The Structure of the Argument

The above argument should by now be familiar; what remains to be shown is the way in which schizophrenia both invites, but also confounds, theorisation by an alienated psychology. Consider the confounding first; on the face of it cognitivism offers what is, intuitively, a highly *plausible* foundation for a psychological explanation of the schizophrenic condition. For schizophrenia after all means a *split mind*, a mind, that is, split not into two or more separate personalities (we do not have to do with dissociative identity - or ‘multiple personality’ - disorder), but a single personality that has *in some sense* become internally fragmented. And furthermore schizophrenia is a paradigmatic psychosis - a condition in which the subject has *in some sense* become out of touch with reality.<sup>113</sup> For both these conditions cognitivism has a ready answer: for the first, cognitivism provides a modular metaphysics of the subject, a view of the mind as being composed of discrete and separable, yet causally interacting, elements. For the second, cognitivism views the subject as isolable from the world yet in causal contact with it in perception and action. What, then, would be more natural than to view psychosis as the failure of these causal relations?

The details of cognitivism also offer the promise of more elaborate explanations. For example, cognitivist metaphysics proposes that inner intentions or goals are causally related to outer actions (in

---

<sup>113</sup> It is the ‘in some sense’s that are important. Cognitivism provides us with one - initially plausible but ultimately unhelpful - approach. Part 4 develops an alternative sketch of the meaning of psychotic ego-fragmentation, taking into account Laing’s view that ‘we cannot give an adequate account of the existential splits [in schizophrenia] unless we can begin from the concept of a [mind as a] unitary whole’.

'output'). And there are many features of schizophrenic speech and action – incoherent speech, negative symptoms (paucity of speech, social withdrawal, apathy etc.) and stereotypies, mannerisms and catatonic behaviour – which might naturally be explained as a failure of goals or intentions to cause actions. Bizarre experiences, hallucinations and delusional perception might be explained as a failure of the causal relations that are purported to obtain in the opposite direction – as a failure in 'input', that is, a failure of stimuli to give rise to appropriate 'internal representations'. Cognitivist epistemology, too, has important offerings. Thought insertion and thought withdrawal, the hearing of voices and passivity experiences, could be understood as a failure in 'inner sense' – a failure in our (so-called) 'introspection' that supposedly allows us to offer reports as to the causes of our actions and on the thoughts and experiences that we find in our own minds. And schizophrenic autism, ideas of reference, suspiciousness, persecutory delusions, incongruity of affect etc. could be theorised as a failure to draw correct *inferences* from the behaviour of others as to the contents of their minds, inferences that are sanctioned by (as they would have it) our folk-psychological 'theory of mind'.

Nevertheless, cognitive accounts of schizophrenia, so it seems to me, inevitably fail to do justice to the psychotic condition. Some of the reasons should already be clear: the second section of this thesis has argued that the cognitivist account of mind is irremediably flawed. In both its metaphysics and its epistemology it fails to account for our mindedness, and so cannot therefore provide the framework for a theory of the breakdown of such.

The further aspects of cognitivism's failure have to be understood in the context of its apparent success vis-à-vis schizophrenia. On the one hand (as I have already argued) *cognitivism provides an alienated conception of everyday rationality and mindedness*, but on the other (as I shall argue in what follows) *cognitive theorists tend to assimilate psychosis to far less serious mental aberrations*. The impression constantly gained from the cognitive accounts is that the psychotic patient has merely made a *mistake* in their reasoning - that their *fundamental contact with reality* is unimpaired - or that anomalies are to be found merely in the *content* and not the total *form* of their experience and thought. In other words, the impression gained is that the person with schizophrenia is not really psychotic, but rather simply in *error* in their negotiation of the world. Theories of delusion, psychotic lack of insight, thought insertion and so on make the person with schizophrenia's troubles out to be merely *surface* phenomena -

problems in the *implementation* of thought in action or speech, for example, and not problems in thought itself. The psychotic core is thereby left unanalysed.

The apparent success of cognitivist theories of schizophrenia is, then - if my theory is right - best explained not by their actual success, but by the too neat dovetailing of an alienated conception of mindedness with a de-alienated conception of psychosis. The claim is not that the cognitivist provides us with an alienated *qua* psychotic conception of normality, and then theorises schizophrenia as simply a condition of normality. If that were the case then the cognitivist account of normality could be abandoned as an account of normality but preserved as the framework for an account of psychosis. Rather the claim is as follows: The cognitivist provides us with a shallow conception of normality - that is, of rationality, sanity, everyday action, thought, perception etc. It is shallow in so far as the true depths of our contact of reality are not theorised but simply presupposed. They are presupposed in so far as the terms in which the cognitivist theorises cannot themselves explain but rather presuppose the functions of a self. This presupposing is of a piece with the claim that the self in cognitivist theories retreats behind the body and then behind the mind. The self, rather than being adequately theorised, becomes the meaning-provider for inner representations, the initiator of inner actions, and the observer of the goings on in the theatre of consciousness. The same terms are then used to characterise psychosis, but the retreated self remains firmly in place. Furthermore a shallow conception of psychosis is employed, and because of this the failure of the cognitive theory can all too easily go un-noticed.

In more detail: a) The disengaged metaphysics provides a framework for a mechanistic account of schizophrenia that locates psychotic disturbance in the dislocation of independently conceived faculties. b) The alienated first-person 'introspective' epistemology allows for a theory of psychotic experience and experience-based delusions as a matter of the misapprehension of the contents of our own minds. c) The alienated representationalist epistemology which grounds our contact with reality in the having of inner representations that accurately mirror the world, or which more generally grounds our contact with the world in representational knowledge (in 'knowing that' rather than 'knowing how'), makes room for an appreciation of the terrible strangeness of delusion as a mere *mismatch* between representation and world or social consensus. But more generally, d) the ultimately tacit and illicit taking-for-granted of the subjectivity and agency of the person by cognitivism (its committing of the 'homunculus fallacy'), and

the parallel fallacy embodied in the computer metaphor (the 'homunculi' here being the illicitly taken-for-granted designers and users of computational artefacts), paves the way for an apparent understanding of the psychotic which depends upon tacitly locating a fully *sane* subject, a rational homunculus, within the insane patient. It is, on this account, only the rhetorical manoeuvres of mentalism which suspend the impression that the analysis of insanity cannot possibly have run deep enough to capture the essence of psychotic ego-fragmentation, an impression which is otherwise all too clear in the tendency of the psychotic subject, as theorised by cognitive theory, to be perfectly sane underneath it all.

It would be possible to use the example of psychosis to argue transcendently against the value of the cognitivist's theorisation of agency and subjectivity. The argument would go as follows: Psychosis embodies a fundamental disintegration of the self and a dislocation of the self from its world. It can be shown that cognitivist theories of psychosis cannot capture this fundamental disintegration. They fail to mine the depths of the psychotic core. This seems to be a pervasive feature of the theories, and not a simple failure in the details. But their theory of cognitive dysfunction is grounded in their theory of cognitive function, so a general failure in the theory of mental illness reveals a fundamental failure in the theory of our integration and world relations. This procedure is not totally without merit: especially when considering delusion (chapter 6) it will I suggest be worthwhile to step outside of the theoretical framework provided by cognitive theories and consider whether delusion as theorised by the cognitive psychologist shares the disturbing features which intuitively and clinically it is known to possess, and whether therefore the cognitive theories are actually *about the right phenomenon*. Generally however, in what follows the critique proceeds by revealing the influence of an alienated conception of mind within the psychopathological theories, and then by detailing the ways in which this reduces the explanatory value of the theories whilst also disguising the weakness in question.

Finally, it is important to bear in mind that the target of the critique in what follows is not cognitive psychology or cognitive theories of mental illness *per se*. There is nothing wrong with investigations of for example the short and long term memory, the comprehension, the responsiveness, the perceptual capacities, the IQ, or what have you of the patient with schizophrenia - although whether fundamental self-disorders could ever be theorised in such terms is another matter. The criticism is aimed not at cognitive psychology but rather the philosophical doctrine of cognitivism; cognitive psychological



theories are only the legitimate target of the criticism developed here when they betray an ineliminable commitment to a cognitivist conception of mind.

## **2. Schizophrenic Motivation and Action**

### **i. Frith's Cognitive Theory of Schizophrenia**

The best-known and most detailed contemporary cognitive account of schizophrenia is that provided by Christopher Frith, most notably in *The Cognitive Neuropsychology of Schizophrenia*. It is this text that will receive the majority of our attention, especially in this chapter (for other cognitive theories of action failure are thin on the ground).

Frith's cognitive account can be broken down into two major components: One part theorises various schizophrenic symptoms, particularly those that have a behavioural rather than experiential character, as disorders of 'willed action'; another views various other symptoms, particularly those with an experiential basis, as disorders of 'self-monitoring'.<sup>114</sup> Disorders of 'willed action' will be considered in this chapter, disorders of 'self-monitoring' in chapter 5.

Within Frith's scheme, action is very much theorised in line with the cognitivist's conception of behaviour as 'output'. Action, that is, becomes *motor response*, its immediate causal precursors are *inner intentions*, and these in turn are *causally* responsive to the other contents of the mind, particularly the *goals* and projects of the subject. This first section considers the 'failure of action' theory, the view that much schizophrenic symptomatology can be understood as a failure of mind to be implemented in action, in its generality. Later sections detail the application of the theory to the topics of *affect* and its expression in gesture, and also to *thought* and its expression in language.

### **ii. Failure of Action: The Theory**

---

<sup>114</sup> A third component involves disorders in monitoring the intentions of others. The final chapter (7) of Frith's book attempts to subsume these three components within an overall theory of schizophrenia as a disorder of 'metarepresentation'. This is arguably the least successful aspect of the theory, and is a project that in any case Frith has since abandoned (personal communication).

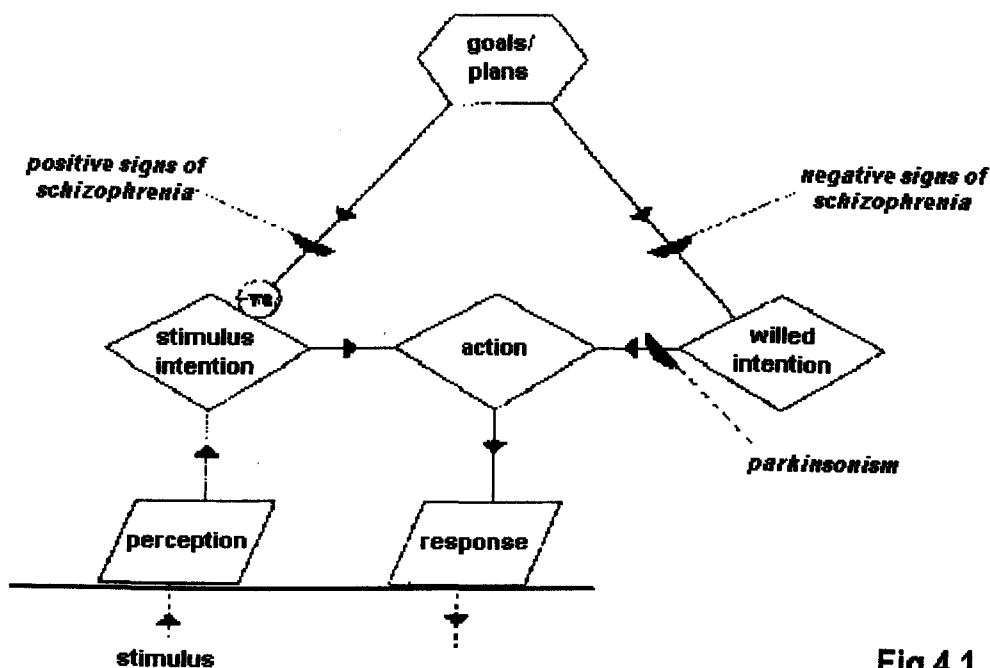
Prominent amongst the behavioural abnormalities found in schizophrenia include the negative signs: poverty of action and poverty of speech, flattening of affect and social withdrawal, and the characteristic positive symptoms: stereotypies, perseverations, incoherence and incongruity of affect.<sup>115</sup> Frith's theory of action failure aims to provide an account of (p.42) 'the cognitive basis' or (p.51) 'the underlying cognitive deficits' of such signs and symptoms.<sup>116</sup> A model is proposed which (p.43) 'assumes that there are two major sources of action. Some actions are carried out directly in response to environmental stimuli. Others are seemingly spontaneous and self-initiated.' The further relevant assumption is that 'patients with behavioural features have a specific difficulty with the latter type of action.' A cognitive diagram is provided to illustrate the difficulties envisaged in the generation of action. (Fig 4.1; p.46).

---

<sup>115</sup> See chapters 4 & 6 of *The Cognitive Neuropsychology...*

<sup>116</sup> Signs are, classically, the observable manifestations of a disease; symptoms, that of which the patient complains. In the present context, signs represent behavioural disturbances, and symptoms experiential disorders. Frith takes over Crow's (1980) distinction between negative and positive symptoms, but instead of understanding these as, respectively, abnormal by their absence or presence, defines positive symptoms as abnormal experiences, and negative symptoms (or rather, *signs*) as abnormal behaviour (p.12).

Frith's use of the sign/symptom distinction is not uncontroversial. Not only is the application of 'symptom' to the manifestations of psychosis unclear. (Consider a patient without insight: it is not that they are complaining of being *ill*.) But the talk of behavioural signs also *suggests* that the observing doctor is merely seeing the outer manifestations of the illness, as if the psychotic core were (as in the mentalist's conception of mind as 'inner') hidden *behind* the behaviour, as if the doctor cannot also just see or hear that the patient is hallucinating or that they are deluded.



**Fig 4.1**

The diagram shows two (of what could be called) cognitive lesions: Positive signs of schizophrenia are explained as a failure by goals or plans to inhibit 'stimulus intentions', which are considered to arise directly in response to environmental stimuli. Negative signs are considered to be caused by a failure of goals and plans to give rise to 'willed intentions'. Various negativities in action, then, are due to an absence of willed intentions, a situation which can be contrasted with Parkinsonism, where action fails to occur even though willed intentions are present.<sup>117</sup>

Although Frith specifies that *cognitive* explanations should be logically independent of *neurophysiological* presentations (pp. 25-27), he doesn't provide *cognitive* criteria for distinguishing self- from other-generated actions. He does however give several examples. Other-generated actions include: yes/no responses to questions (p. 43), answers to questions in which the required type of response is indicated by experimenter, reactions which seem appropriate in the light of responses to similar previous actions (p. 50), and a subject's manipulation – experienced as unwilling – of ready-to-hand objects, as caused by direct stimulation of the cingulate cortex (p. 54).

<sup>117</sup> That is, the cognitive lesion occurs in a different spot. See Fig. 4.1.

The scope of the model is further indicated by what it suggests about the subject's actions. For example, Frith asks (p. 45):

What will happen if you cannot generate a spontaneous new response [in, e.g., a fluency task]? There are three possibilities. First, you might do nothing (poverty of action). Second, you might repeat your previous response, even though it is now inappropriate (perseverative, stereotyped responding). Third, you might respond inappropriately to some signal in the environment (stimulus-driven behaviour...)

So a subject asked to list as many names of animals as they could off the top of their head in three minutes (Example 4.3 p. 47) either produced very few animal names, the same animal name several times, or a list containing the names of several non-animals. Incoherent and incongruous responses (such as the non-animal responses) are theorised as due to an excessive determination of action by irrelevant stimuli (caused by a failure to inhibit 'stimulus intentions').

### **iii. The Explanatory Power of Frith's Action Model**

*Within* the context of the theory, the content of the distinctions and hypotheses seems clear. Stimulus intentions are those intentions caused by the environment, whilst willed intentions are self-generated. Willed actions are those actions caused by willed intentions, stimulus-driven actions are those caused by stimulus intentions. The negativities in schizophrenic failure of action are a function of impairments in the generation of willed intentions; the positive symptoms stem from a failure in the negative feedback on stimulus intentions.

Being able to spell out the theory is not however the same thing as understanding it; the difficulties arise when we ask what the theory *means*. Consider the negative signs first. The theory presents itself as a genuine explanation of why the negative signs occur. They occur because there aren't the willed intentions around to cause actions to happen. The rhetoric of the theory (and the details of Figure 4.1) suggests both that intentions are internal phenomena which can give rise to actions so long as impairments in the causal pathways are not present, and that goals or plans are internal mental phenomena which sit further back in the causal system.

If the theory could only be understood at this level, and if the arguments against the cognitivist conception of the relation between goals or intentions and action presented in chapter 3 were along the right lines, the theory can be seen to be worthless *ab initio*. For both goals and intentions are not (unless we are being secretive) characteristically hidden behind action; they are rather imminent within the behaviour itself. This can be seen, incidentally, in the in-aptness of describing Parkinsonism as due to a failure of willed intentions to give rise to action. For whilst the patient with Parkinson's might *want* to be able to pour and drink a glass of water, if their shaking completely and predictably incapacitates them in this, it would be stretching the concept of 'intention' to suggest that they intend to do so, even when they give it their best shot. If intentions and plans are characteristically imminent within action itself – and when we are concerned with everyday rather than premeditated action they typically are so imminent – then failures in action cannot be theorised as due to a failure of goals to 'give rise to' intentions or a failure of intentions to give rise to actions, for we do not have separate items to stand in such external causal relations in the first place.

If scientific theories are to be understood as necessarily having certain tacit metaphysical or epistemological commitments, and if such commitments are questionable, the resulting theory will not stand a chance. This however is not the only way to read such theories. What needs to be seen is what explanatory and predictive power the theories have when they are translated out of the cognitivist rhetoric.

If the disengaged conception of action is abandoned, Frith's theory can be viewed most simply not as a causal explanation but as a phenomenological description. The theory becomes not an explanation that certain movements fail to occur in a subject suffering from schizophrenia because their goals fail to give rise to willed intentions. Rather we are given a description: intentional action in the schizophrenic is unlikely to occur. But no explanation for this is offered. In this sense the effect of the cognitivist rhetoric is not simply to invalidate the theory *ab initio* by implicating within it a hopeless metaphysic. Rather the rhetoric serves simply to disguise the lack of explanatory prowess possessed by the theory; it gives rise to a mere appearance of explanatory power and a mere appearance that the psychological phenomena can receive an apt theoretical treatment with the methods found in the natural sciences.

To leave the discussion at this point would however be unfair, for the content of Frith's claim is in fact considerably more specific, as can be seen by means of the comparisons Frith draws between the schizophrenic symptom and other disorders. The patient with Parkinson's disease also has 'negative symptoms' - they also experience a failure of action. But the problem in this case is not a problem in forming intentions to act, but rather (p. 55) 'a difficulty at the stage of motor output'. The patient with schizophrenia, by contrast, 'probably has no action in mind to perform'. Further content now comes from a comparison with dementia: it is not that the patient has no action in mind because they have lost their intelligence or their memory, and neither is it the case that they have an absence of 'goals and plans'. A final contrast with depression finishes the clinical picture (p. 48): the depressed patient *doesn't* want to act, but the schizophrenic subject *cannot* want to act.

The contrast with Parkinsonism is not perhaps as clear as one might like, since the requisite contrast in ascription conditions for goals/plans and intentions is not totally obvious: in the context of answering a question or performing any typical unpremeditated task the distinction between having an intention and having a goal is not always easy to draw. But whatever the difficulties the central content of the distinction is just that between a *mental disorder* and a *motor disorder*: the mind of the subject with Parkinsonism remains intact, they merely experience a difficulty in movement. The patient with schizophrenia however has no difficulty in movement *per se*, but rather a difficulty in *intentional action* - a concept falling in a psychological rather than a physiological category. The contrast with depression is again somewhat unclear: the ascription conditions for being unable to want to act (the supposed predicament of the subject with schizophrenia) are not transparent. But once again the real content of the distinction lies in distinguishing a *severe* (psychotic) mental disorder from a *less severe* (neurotic or non-psychotic) mental disorder. The depressed person's failure to act is folk-psychologically intelligible<sup>118</sup>: they don't act because they don't want to (something with which we are all familiar - although we might not understand why they don't want to act). The psychotic patient's failure to act is not intelligible in *this* way: they do not act because they 'cannot' want to, a failure which is not folk-psychologically intelligible.

---

<sup>118</sup> For details of the use of this criterion for psychosis by Jaspers see chapter 7.

The distinction between the psychological and the physiological is manifest not only in the distinction between the neurological disorder of Parkinsonism and the mental illness of schizophrenia, but also in the distinction between *willed* intentions and actions and *stimulus-driven* intentions and actions (see Figure 4.1). The nature of a 'stimulus intention' is never precisely spelled out, but what seem to be in question are the kind of motor programs that are activated by familiar objects. I see a doorknob and make for to grasp it; I get on my bike and automatically adjust my posture in subtle ways to compensate for various instabilities encountered in cycling along. The level of description of such phenomena is that of the automatic *response* or *behavioural subroutine*: they do not constitute full-blown intentional actions. As shall be seen below, Frith accounts for various positive symptoms in terms of a failure of negative feedback on 'stimulus intentions', but what is important for now is just to note that in the theory (and in Figure 4.1) the sub-personal 'stimulus intentions' and the fully personal 'willed intentions' are positioned *not* (as one might expect) on *different logical levels*, with intentional actions being emergent psychological phenomena out of a lower logical level of behavioural subroutines under various degrees of feedback control, but rather (as cognitivism would prescribe) simply *side by side*, as phenomena in the *same logical class*. This conflation of logical levels is encouraged by a cognitivist metaphysics, with its insistence that intentions are to be viewed as inner states causally operative in bringing about action.<sup>119</sup>

In some cases the conflation may be harmless. But in the present instance the cognitivist's interiorised conception of the mind leads directly to a serious theoretical problem. The claims of the cognitive theory - the failure of inner 'willed intentions' to give rise to actions - ask to be read as causal hypotheses, and hence as supplying genuine causal explanations of the behavioural phenomena in question. But what in fact is achieved is something considerably less impressive (although not worthless): the confirmation of schizophrenia as a genuinely psychological and severe (i.e. psychotic) disorder, and the confirmation of the fact that it involves a failure in intentional action. On the one hand these are not facts previously unknown, though this does not mean it is unhelpful to repeat them (although the cognitivist idioms and the appearance of genuine explanation that they encourage are not so

---

<sup>119</sup> It is also not I think insignificant that Figure 4.1 has actions - a psychological (and hence on a cognitivist story, internal) phenomenon - as responsible for the causation of responses - a more primitive physiological phenomenon.

helpful). On the other, it is not I think insignificant that the theoretical framework employed is that of the cognitive study of neurological disorders. What is found - although this is not of course how it is stated - is that schizophrenia cannot be theorised in the normal cognitive terms - in terms of behavioural subroutines, motor failures, feedback and control etc. - but must be situated within a properly *intentional* context. Of course on the one hand this is news to no-one (excepting the cognitive psychologist as yet unfamiliar with the ascription conditions for mental illness and the categorical structure of the conceptual network in play), but on the other it highlights, by means of a kind of 'differential diagnosis', an essential condition for talk of psychosis.

To sum up: Frith's theory presents itself as a causal explanation as to why the negativities in action experienced by the subject with schizophrenia occur. In fact, however, unless the cognitivist theoretical framework is left in place - in which case the theory simply fails due to its implausible metaphysical commitments - the theory has only a descriptive rather than an explanatory value. The subject with schizophrenia becomes incapable of intentional action, in contrast with the depressed subject who simply doesn't want to act, with the subject with Parkinsonism who might try to act and fails, and also with the demented subject who has no understanding left on which to act. The normal sub-personal ingredients of cognitive models (albeit with their personal level names) of stimulus 'intentions' and various motor programmes and behavioural subroutines are joined together side by side (on the same logical level) with personal level notions such as goals, plans and willed intentions. This is encouraged by the cognitivist metaphysics which views all such phenomena as self-contained inner mental events and states causally operative in the bringing about of behaviour or other such events and states. What this in turn encourages is the view that psychological disorders - disorders which (whatever their aetiology) by definition have to be understood as disorders of the person (in action and not movement, in thought and not information processing) - are capable of being theorised within the framework of cognitive neuropsychology. Once the logical levels have been separated out, however, what the model clearly demonstrates is the inability of schizophrenic action failure to be theorised in such terms.

#### **iv. The Predictive Scope of Frith's Action Theory**



So far cause has been found to reject both a metaphysically underwritten and a cognitive neurological reduction of the psychological theory. Given that a considerable portion (pp. 53-63 and 108-112) of Frith's text is given over to detailing possible neurological underpinnings of schizophrenic action failure, this latter rejection may appear precipitate. Recasting the theory in a purely neurological idiom would not however be of benefit, for what a cognitive *neuropsychology* of schizophrenia aims to provide is a *cognitive psychological* characterisation of the disorder; only when this is adequately provided can the characterised phenomena be neurologically explained. So long as Frith's theory is allowed to define its own objectives, so long as a psychological understanding of schizophrenia is being sought, the neurological reduction must be avoided.

To provide a final assessment of the cognitive theory of schizophrenic action failure consider again the fluency task mentioned at the end of section ii. above. Whatever the metaphysical credibility of the theory, its empirical credibility and explanatory validity – one might think – is guaranteed by the predictions it can make. This is what Frith says (pp.46-7):

My model for poverty of action proposes that schizophrenic patients with negative signs have difficulty in generating actions spontaneously. On the basis of this model we would expect such patients not only to show a lack of action. In certain circumstances we would also expect them to show stereotyped behaviour or an excess of stimulus-driven behaviour. Thus the same underlying deficit can lead to different kinds of surface behaviour.

Furthermore, the same surface behaviour can be accounted for by different underlying deficits (p. 48). For example, both a schizophrenic patient and a demented patient may respond with few words in a fluency task. But:

Schizophrenics with negative symptoms produce few items because they find it difficult to perform a self-directed search. Demented subjects produce few items because the inner lexicon itself has become depleted and contains fewer words. Thus, however efficient the search, only few items will be found.

Translate out the unnecessary cognitivist rhetoric, of 'inner lexicons' and 'self-directed searches' etc., and what we have is a fully intelligible and testable hypothesis: Demented patients would not even

in other situations be able to provide the names of, for example, many different animals. If they were presented with photographs of a trip to the zoo, they wouldn't be able to name many of the animals, whereas a schizophrenic patient should have no trouble with this task. The same surface behaviour can be fitted into a different broader pattern of behaviours and capacities. And this shows the potency of the cognitive theory for accounting for some particular instance of symptomatology.

But how powerful is this model in the present instance? It is hardly news that the schizophrenic patient is not demented – hence the well-known inappropriateness of the 'dementia praecox' diagnosis. This aspect of the model – the differential explanation of the same 'surface' behaviours - makes no novel predictions and does not explain what is already known; it merely recasts the description in a rather unhelpful cognitivist idiom.

Consider though the different 'surface' behaviours that may result from the same underlying impediment. Frith asks (to quote the passage again):

What will happen if you can not generate a spontaneous new response? There are three possibilities. First, you might do nothing (poverty of action). Second, you might repeat your previous response, even though it is now inappropriate (perseverative, stereotyped responding). Third, you might respond inappropriately to some signal in the environment (stimulus-driven behaviour, or what Luria (1973) calls an inert stereotype).

Well, if we cannot 'generate a new response', if, for whatever reason, we can't think of anything appropriate to say, we may well say nothing. And true, we may also repeat a previous response. But this latter outcome is no more predicted by the theory than the possibility that we may, say, stand on our head or go brush our teeth. It is a *logical* possibility, true, but not one that the theory actually predicts – only one that it fails to rule out. The same obtains as well for the third possibility - of responding inappropriately. Neither of these possibilities is actively predicted by the suggestion that the schizophrenic cannot 'generate a spontaneous new response'.

If we return to the cognitive model which demonstrates the theory, what we in fact find is that the different possibilities of response are *not in any case* to be explained by 'the same underlying deficit'. The positive signs are a function of one cognitive failure: a failure of goals and plans to inhibit stimulus-driven actions, and the negative signs are a function of a different cognitive failure: a failure of goals and

plans to give rise to willed intentions. (This is shown clearly in Figure 4.1.) Later in the book a rationale for the conjunction of the two cognitive failures is somewhat incidentally given (p. 53):

... (2) the link between goals and actions is necessary, not only for the initiation of acts, but also for the termination of actions, as actions are normally terminated when the goal is achieved. Lack of this normal termination results in perseverative and stereotyped behaviour; (3) the same mechanism that initiates and terminates actions, also inhibits inappropriate stimulus driven actions. Lack of this inhibition leads to incoherent behaviour.

What to do with this rationale is, however, another matter. To be sure we must be able to tell when we have achieved what we set out to do, otherwise (unless we get bored or tired) we will not stop trying to do it. But why this should in some sense be the same function as our acting in accord with our goals is unclear. If the mechanistic rhetoric is taken at face-value, we might have the appearance of an explanation, but then the theory presents no evidence that the same mechanism that initiates and terminates action should also inhibit inappropriate stimulus driven actions. Psychologically speaking, what we have been given is a random hodgepodge of descriptions of tendencies; neurologically speaking, what we seem to have is an example of what used to disparagingly be known as 'brain mythology': an entirely speculative suggestion about some as-yet-unidentified brain structure. The cognitivist's mechanistic rhetoric may appear to bridge the two levels of description, but in reality it simply conflates them, leaving us without a coherent psychological explanation (and an as-yet only speculative neurological explanation) of the behavioural abnormalities of schizophrenia.

### **3. Schizophrenic Affect and Expression**

It is in particular Frith's theoretical account of the *emotional* disturbances of the schizophrenic that demonstrates how an implicit cognitivist metaphysic can conspire with and render implausibly plausible an alienated phenomenology of mind, and thereby both inspire and generate a perceived need for a cognitive psychological theory.

## i. Frith's Cognitive Theory of Emotion

*Incongruity of affect* is theorised by Frith (p. 50) as reflecting the third cognitive dysfunction mentioned above, that is, as caused by the patient's being 'captured' by immediate aspects of stimuli rather than responding to the total emotional structure of the social situation. This is surely an empirically plausible theory, and represents a genuinely psychological explanation: a particular set of behaviours being understood as instancing a general disposition (or in the cognitivist idiom: various 'surface' behaviours being 'caused' by the same 'underlying' deficit.) True, construing incongruity of affect as merely a 'behavioural sign' may appear to evince the cognitivist's dichotomising of thought and feeling from their bodily expressions, locating the disturbance merely within the latter sphere, but this is incidental when the genuine content of the theory is considered. No evidence however is provided for the theory, and it will not be touched on further here, although the general topic will be re-addressed in chapter 7.

*Flattening of affect* is similarly labelled a behavioural sign, and a consideration of the diagnostic situation in which flattened affect is assessed is used to suggest that (p.51) "flattening of affect" could be relabelled as "poverty of [emotional] gesture". The psychiatrist, that is, makes the judgement that a subject suffers such emotional blunting by noticing the absence of the subtle shifts of expression – laughter, smiling, scowling – that generally accompany interpersonal communication:

We laugh or make a sad face to indicate that we are being facetious or that we regret having to be critical. Even a stiff upper lip or poker face can be put into the service of deliberately communicating an attitude (e.g. stoicism, mistrust). It is impairments in producing these subtle aspects of non-verbal communication that are rated as "flattening of affect".

What is normally understood as a lack of emotion is, then, reinterpreted as a problem in the mechanics of emotional expression. It 'can more readily be seen as another example of a lack of spontaneous, self-initiated action'; more readily, that is – at least, this is the implication – than as a genuine deficiency in feeling itself. This is explicitly stated later in the book (p.102): 'I consider that the

sign “flattening of affect” does not actually refer to affect, but to a lack of expressive use of the face and tone of voice in communication.’

Whilst the general form of psychological explanation on offer here is perfectly respectable, one of subsuming some particular set of symptoms within a general pattern (i.e. an explanation by generalising redescription, the details are suspect on phenomenological, psychiatric and conceptual grounds. These shall be considered in turn.

## ii. Critique of the Cognitive Theory of Emotion

First, the phenomenological description of our everyday expression of affect is askew. We may *sometimes* put emotional expressions on our faces to convey some socially pertinent signal. And *occasionally* too we might attempt to deliberately communicate some attitude of mistrust by means of a poker face. But most of the time our emotional reactions are spontaneous and non-deliberative. Our affective expressions put *themselves* on our faces, and are not put there *by us*. However responsive emotional expression can be to social situations<sup>120</sup>, we are usually unconscious of our miens, our attention focussed on the people we are with or on the topic at hand. Putting it simply: emotional expressions rarely count as actions, and are rarely preceded by intentions, and it is hard therefore to understand an absence of such expression as a failure of intention in action.

Second, in order to subsume flattening of affect within the category of failures of action or even within a failure of expression, Frith is forced to re-describe the clinical situation and subvert the standard clinical impression that what is missing in the schizophrenic is affect itself. The cognitive account can be compared and contrasted with any of those provided in the psychiatric, psychological or psychoanalytic literature. Here is one example, from Arnold Buss’s *Psychopathology* (p. 197):

The major dimension of affect is the one that ranges from elation to depression, and it is here that the schizophrenic suffers a basic inadequacy. There is a relative absence of joy or sadness that is called “blunting of

---

<sup>120</sup> Consider that many people laugh and smile far more at an amusing object or film when in company than when alone. This fact alone, however, does not justify an analysis of expression as in essence performing a *purely social function*; even less so does it sanction an account of expression as *deliberative gesture*.

affect.” A schizophrenic may describe harrowing life experiences or the details of a somatic delusion (“my insides are rotting away”) without displaying signs of worry or melancholy. This emotional apathy is seen most clearly in the later stages of schizophrenia. In the early stages of the psychosis, the schizophrenic may still have some emotional involvement, some degree of appropriate affective response to the ups and downs of everyday life. The early schizophrenic can still experience some joy over good fortune, some sadness over misfortune, but gradually these mood reactions disappear. ... It is not so much that life depresses the schizophrenic but that it leaves him apathetic, and it is believed that this emotional dullness is present from the very beginning in schizophrenia.

Suffice it to say that the application of Frith’s scheme requires considerably more justification than he provides if it is genuinely to overturn a century of fairly unanimous psychiatric observation<sup>121</sup>. This is not to doubt the truism that the psychiatrist assesses the emotional state of the patient by looking at and listening to how they feel. This is the case with any assessment of emotion, but it does not mean that we are not thereby in contact with the emotions themselves, that we can only *actually* assess the ‘outer signs’ of the ‘inner’ feelings.

This leads us, thirdly, into a consideration of the conceptual difficulties of the cognitive theory. On a cognitivist reading of the nature of mind, it will come as no surprise that in psychopathology emotions may fail to issue their standard expressions. For emotions are conceived by the cognitivist as inner states which are causally responsible for the distortions of countenance and inflections of voice by which they are publicly advertised. Whether or not some particular emotion gives rise to some particular expression is a purely contingent affair, depending on the operation of the requisite cognitive mechanisms mediating between the inner mind and the outer body.

Given this cognitivist reading, the content of Frith’s theory is easy to comprehend: the schizophrenic subject may well have emotions, but the cognitive mechanisms by which the subject generates willed intentions to act emotionally are all awry.

If, on the other hand, we reject this conception of the nature of mind as irremediably flawed, as was recommended in Part 2 of this dissertation, then the cognitive theory must be re-thought. For on the view offered there, emotions are not inner states of the organism merely contingently related to behaviour. They are dispositional in character, and have their own characteristic expressions, expressions which are

---

<sup>121</sup> Though see Louis Sass’ *Schizophrenia, Self-Experience, and the So-Called “Negative Symptoms”*.

defeasible criteria - and not merely contingent evidence - for the emotions themselves. Smiling is part of the natural repertoire of happiness: being happy is, amongst many other things, being disposed to smile. It is because of this criteriological connection between emotion and expression that a psychiatrist, for example, is not left out of contact with the emotions themselves when carrying out a present mental state examination. They are not restricted to observing *mere* behaviour, for in seeing or hearing the behaviours they are in epistemic contact with the emotions themselves. If, then, the psychiatrist finds themselves in a situation when a patient is acting unemotionally, then they are entitled to assert, without any intervening inference, that that is indeed how it is with the patient. This, then, is the diametrically opposed point of view from that lurking behind the traditional 'problem of other minds'. What is required is not a piece of reasoning to take us from observed behaviour to hidden mental state, but rather a piece of reasoning explaining why, in some given instance, an emotion ascription (or, say, an ascription of affective blunting) should be *withheld*.<sup>122</sup>

In his theory, Frith provides no evidence that the schizophrenic subject is after all affectively intact and that their problems reside solely within their capacity to express themselves in action. Given the dubious character of cognitivism, the burden of proof must be on such a theory to explain why an ascription of affectivity is not withheld in the absence of affective behaviour. This issue of the burden of proof is taken further in the following section on thought and language.

#### 4. Schizophrenic Thought and Language

##### i. Incoherence of Speech

*Incoherence of speech* (p. 50) is, as with incongruity of affect, modelled by Frith as a matter of inappropriate responses elicited by immediate details of stimuli. The cognitive theory explains this as due to the absence of negative feedback on 'stimulus intentions' from goals and plans (c.f. Fig 4.1). As with the theory of emotional incongruity, the presumption that incoherent speech reflects (or at least, *often* reflects) capture by irrelevant details of the communicational or visual scene is fairly uncontroversial.

---

<sup>122</sup> C.f. chapter 3 subsection 2.

Whether or not this can be understood non-metaphorically by reference to a failure of negative feedback on stimulus intentions from goals and plans is more controversial, as is the explanatory purpose of such a redescription.

There is of course a *normative* connection between a goal and its realisation. If we are to follow some goal, then we are not to make irrelevant responses, for such do not count as following the goal. But what this testifies to is an *internal* relation between a goal and its realisation, not an *external* (causal) relation between an inner state and an outer process. Cognitivism's mechanistic conception of thought encourages the conflation of these relations, a conflation which, once made, naturally sustains the presumption that there just *must* be some kind of negative feedback from the goal on stimulus-driven responses.

Talk of negative feedback has an established place within a *systems* discourse, naturally applicable to organisms and machines<sup>123</sup>; but the principles for an extension of this into a discourse of the person are unclear. How could it be determined, for example, whether the chaotic ascendancy of sub-routines of object-response in the schizophrenic is a result of a failure in some normal psychological inhibition, or whether it is a matter of simple over-activation, or whether we have here to do with an absence of a positively organising force from higher psychological functions?

This latter question appears to have an answer within the cognitive theory. Figure 4.1 shows a situation in which intact goals and plans fail to suppress stimulus intentions, resulting in some of the positive features of schizophrenia (incoherence in action). But whilst the cognitive theory is surely reasonable to explain the abnormal verbal fluency of a person suffering from schizophrenia in terms not of a failure in aims – given the avowals of intent and the apparent effort in pursuit of this intent – but rather in a failure in effecting these goals, it is still unclear how the aims themselves act to inhibit automatic responses. Furthermore, the capacity of the schizophrenic to sustain the attribution of the goals may in the relatively simple context of the verbal fluency experiments be relatively uncontroversial, even when responses are produced that might in different circumstances contraindicate the attribution. But such experiments are not real life: we are not normally given by someone else a simple task of listing as many animals as we can off the top of our heads in three minutes.

---

<sup>123</sup> C.f. Ilya Prigogine, *Order out of Chaos* & Stuart Shanker, *Computer Vision or Mechanist Myopia?*



In the normal run of things our goals, such as they are, *are far more imminent within our actions themselves*, to be read out of this context, and not supplied 'externally' by an experimenter. Half way through a conversation I start to tell you about my holiday, about the beaches and the food, the hills and the wild flowers. Perhaps after a couple of minutes the conversation turns to general botanical topics. Later in the day a friend with schizophrenia starts to tell me about their holiday, but after mentioning the beaches they start to ramble about peaches and leeches, and talk of some flower leads into a discussion of baking. To theorise the latter as a lack of negative feedback from conversation goals now seems an arbitrary imposition, given the irrelevance of goal attributions in the first conversation, at least in so far as goals are considered as pre-dating the actions that manifest them. We do not have to do here with such goals, the attribution of goals simply being based on the conversational content and structure itself. The latter conversation to be sure lacks structure, but it is just this very lack of structure that precludes the attribution of those goals that, on the cognitive theory, are to be considered present and intact but without the capacity to inhibit alliterative rambling and so on.<sup>124</sup>

## ii. Thought Disorder

The somewhat unhelpful influence of cognitivism on the psychological modelling is even more apparent when attention is turned to the schizophrenic symptom of *thought disorder*. This influence is so striking that it is worth quoting in full the passage in Frith's *Cognitive Neuropsychology...* that introduces the topic (p.97):

The peculiar speech observed in many schizophrenic patients is traditionally labelled "thought disorder".

This label suggests that the peculiar things that schizophrenic patients say are a consequence of peculiar thoughts.

---

<sup>124</sup> What this surely demonstrates is the complicity with which the procedures of experimental psychology enter into alliance with the disengaged theorizing of cognitive psychology. By specifying a highly determinate context which involves specific tasks, goals, plans and *procedures*, we are lulled into the supposition that the cognitivist's conception of intelligent action as *guided action*, action guided by pre-existing intentions, is universally applicable. Whereas a phenomenology of everyday cognition reveals a quite different picture: of intention and goal as solely *imminent within action*. (C.f. Stuart Shanker's *Wittgenstein's Remarks on the Foundations of AI*.)

The label further suggests that the ability to put these thoughts into language is unimpaired. So far this assumption remains unproven. Indeed, first-person accounts suggest that some patients at least do experience difficulty in putting their thoughts into language.

Frequently, patients express abnormal thoughts in normal language. Hence the expression of the false beliefs associated with delusions can be understood as a consequence of abnormal thought processes. For instance, one patient told me, "the reason I get sunburn is because people are lying under sun-ray lamps and thinking about me." Psychiatrists therefore distinguish between disturbances of the content and the form of thought.

If a patient has "formal thought disorder", then it is not necessarily the content of their thoughts that is abnormal. It may be the form in which the thoughts are expressed that is abnormal. In this case there are abnormalities in the language used to express the thoughts. There is a fundamental difference between language and thought, which has received surprisingly little emphasis in the study of schizophrenia. Thinking is a private matter, whereas language is arguably the most important method we have for communicating with others. Thus language is not simply the expression of thoughts; it is the expression of thoughts in a manner designed to communicate these thoughts to others.

This idea, that what psychiatrists have traditionally understood as thought disorder may in fact involve a difficulty in putting intact thoughts into language, allows the symptom to be encompassed within the general 'failure of willed action' scheme.

There is no doubt of course (to consider the quoted section: '[There is a fundamental difference between language and thought, which has received surprisingly little emphasis in the study of schizophrenia...'] that 'thought' and 'language' are fundamentally different concepts. Thinking is something we do, a thought something that we have, but a language is not something we do and in as much as we have a language (English or French, say) this is not an occurrent function of our minds but a general capacity of our person. But whether this difference is adequately characterised by a private/public distinction is debatable. What Frith seems to have in mind is the possibility that we can keep our thoughts to ourselves. But then we can also use language to talk to ourselves in our heads, whatever its possible use in communication. And we frequently let our thoughts be known to others.

More importantly, even if 'thought' and 'language' are different concepts, this does not preclude the possibility that any kind of complex thought, including most of the thoughts that are attributable to linguistic beings, are intrinsically linguistic. They can only be expressed in language and, arguably, are only ascribable to language-users, since the only way of individuating the content of such thoughts is

through their linguistic expression. Thought, as the well-known Kantian argument goes, is conceptually structured; to think a thought we must possess the concepts which structure the thought; and to possess a concept (of 'thought', say) is to have the know-how required for using the term that denotes the concept (to know the meaning of 'thought' or some translation).<sup>125</sup>

Furthermore, the distinction between form and content employed by psychiatrists is not perspicuously represented as a distinction between the linguistic or syntactic structure of a thought or sentence and the coherence of the topic or the semantics of a thought or sentence. A perusal of the textbooks provides no particularly clear or consistent explanation of the psychiatric form/content distinction, but reveals that the form/content distinction also operates in descriptions of hallucinations, and that 'form' pertains more to the logical coherence of the thought than to the well-formedness of its expression.

Most importantly, it is clear from the description of thought as a 'private matter' that Frith is thinking of thought *qua* inner monologue, an inner process independent from communication. It is just this conception of thought that sits at the heart of the cognitivist's account of cognition: thought as an inner process which is introspected and then reported in language when communication is desired. And it is just this conception of thought as inner process that was argued in Part 2 to be peripheral to our understanding of mindedness, and this epistemological conception of 'first person access' that was argued to be fundamentally misguided. For in the main, the thought that is manifest in discourse is not to be understood as some pre-existing already-elaborated inner structure reported in communication. It is imminent within the communication itself, predicated of the speaker purely and directly on the basis of the sensibleness of the utterance. (To put it paradoxically: *if* it were insisted that thought must be some kind of *process*, then the thought must be identified with the communicative endeavour itself.<sup>126</sup>)

Although it may sometimes be apt to think of thinking as an inner process, a rumination in private soliloquy that may or may not be followed by conversation, it is clear that the thought that the psychiatrist wishes to assess does not fall into this category. The diagnosis of thought disorder is made

---

<sup>125</sup> C.f. *PI* §502, *BB* pp. 4-5.

<sup>126</sup> C.f. Wittgenstein's remark on willing put forward in *PI* §615, for consideration and not for endorsement: "'Willing, if it is not to be a sort of wishing, must be the action itself. It cannot be allowed to stop anywhere short of the action.'"

directly on the basis of an assessment of the psychotic subject's discourse: the disordered thought is directly manifest in the language which shares its structure and constitutes the criteria of its identity. In this respect, the distinction between thought and language on which Frith remarks is not to the point. And the possibility of theorising schizophrenic thought disorder as a disorder of action, of the implementation of thought, is similarly not apt. The thought we are dealing with here is not pre-existing cogitation which must be put into language. It is rather the capacity to be considered a rational animal, the ability to discuss intelligently, to follow through a train of thought or argument in accord with the canons of practical reasoning and logic, to keep to the point. Whatever the difficulties schizophrenic patients may report in putting their thoughts into language, it is scarcely an 'unproved assumption' that disordered discourse is indicative of disordered thought. Given the criterial relation between the two, the burden of proof is quite the opposite of what is suggested: positive grounds would need to be provided if the diagnosis of thought disorder were not to be automatically made on the basis of rationally deviant discourse.<sup>127</sup>

It might be considered that such grounds are provided by what Frith goes on to say about the nature of the communicative deficit which manifests (what is diagnosed as) thought disorder. For he notes (p. 98) that the defects in language are fundamentally 'expressive' rather than 'receptive'; that is, they indicate problems not so much in the understanding of what is being said by others, but in the production of readily intelligible utterances. Thought disordered patients in particular fail to provide adequate referents for pronouns and cohesive ties between ideas. Frith concludes (p. 100) that 'some schizophrenic "thought disorder" reflects a disorder of communication, caused in part by a failure of the patient to take

---

<sup>127</sup> This conceptually implausible dichotomisation of thought and language is not restricted to cognitive theorising about schizophrenia. The psychiatrist Tim Crow - in his paper *Nuclear Schizophrenic Symptoms as a Window on the Relationship between Thought and Speech* - makes the proposal that schizophrenia can be understood as a disorder of hemispheric integration, the dislocation being between the 'signifieds' or meanings in the non-dominant hemisphere and the 'signifiers' or 'phonological output' in the dominant hemisphere. Aside from the unclarity involved in locating 'meanings' within the brain, the theory clearly depends on a conception of meaning as intrinsically non-linguistic, as something to be brought into causal connection with the intrinsically un-meaningful 'sound patterns' of language. Both the non-praxical conception of meaning and the behaviouristic conception of language ought to be questioned; once again, as with the relation of intention to action, of perception to its objects, of a desire to its intentional content and so on: no amount of causal glue can adequately rejoin what has been conceptually sundered.

account of the listener's knowledge in formulating their speech. ... The normal speaker takes account of the listener's lack of knowledge, and thus the schizophrenic listener can understand. The schizophrenic speaker does *not* take account of the listener's lack of knowledge, and thus the listener has difficulty in understanding.'

There are of course questions about whether the subject with schizophrenia's 'receptive' linguistic capacities really are always intact. One criterion, for example, for having understood what someone is saying is whether one could paraphrase what has been said, and this clearly draws on 'expressive' linguistic capacities. In other words, it isn't clear that a *principled* distinction can always be drawn between the two sets of capacities. Nevertheless the phenomenological observation concerning various linguistic capacities is important, even if it is just that: an important part of the characterisation of how subjects suffering from schizophrenia think, of the nature of schizophrenic incapacity; an explanation by re-description, that is, and not a mechanical account of the underlying structure of the deficit.

A similar point could be made concerning schizophrenic failure to take account of the knowledge of the listener when in discussion: for example, the failure to supply the referents of pronouns. The theory that thought disorder really reflects *merely* a disorder in communication and not a disorder in the structure of thought *per se* must presuppose a conception of the structure of thought that excludes as constitutive of that structure the considerateness of the speaker. But this is once again debatable: taking account of the knowledge and understanding of others as a *general* ability is not a completely separate function from being able to formulate a thought in the first place. If we think that it is, then we are doubtless thinking of those situations in which, failing to take account of the lack of knowledge of others, we fail to make ourselves understood. Here we can clarify what we mean, can supply the referents to make the pronouns intelligible; we apologise for our opacity and rectify the situation. This is not the case with the schizophrenic: they do not merely make a *mistake*, failing to bear in mind the level of understanding or degree of knowledge of the listener, but they suffer from a far more fundamental disorder of relatedness. If they were to rephrase what they said, similar problems would arise - and if they did not arise, they would clearly not have been characterised as psychotic in the first place. At this fundamental level of relatedness the possibility of separating out in a modular fashion communicative from cognitive capacities is far less plausible.

### iii. Conclusion

This impression that the cognitive theories of schizophrenia often give, of reframing psychosis as a merely cognitive *error*, is not perhaps surprising. For what it instances is that general trait of cognitivism: the retreat of the true self deep within the subject, forming an illicit and tacit backdrop against which the cognitive theory both *requires* and *refuses* to be understood: *Refuses* to be understood because cognitivism is supposed to have provided a reductive explication of the self, not to have presupposed it. *Requires* to be understood because the modular functions into which the cognitivist analyses the mind cannot be understood for what they are, as they would have it, in isolation from one another, but rather presuppose the general mindedness of the subject, including their agency. This impression of mere error will be considered in further detail in chapter 6 when cognitive theories of delusion are interrogated. To pre-empt the discussion, what is beginning to emerge is that the hidden 'homuncular' true self of the cognitive theories is a fully *sane* subject; that it is this self that is prone to the making of simple mistakes, or failing to put intact, rational, sane thoughts into coherent action. And that it is this that accounts for the unmistakable impression that the cognitive psychologist's subjects are not truly psychotic, that the analysis of schizophrenia has not been pushed *deep* enough, and correlatively that the explanatory power of the cognitive theories vis-à-vis psychosis is minimal.

To conclude, consider how this tendency manifests itself in the various aspects of the cognitive theory of schizophrenic symptomatology as action failure, dependent as it is on a cognitivist's metaphysical conception of mind as essentially non-agential and of action or behaviour as essentially mindless. The initial presentation of emotional expression and thoughtful communication that cognitivism provides, and that the cognitive theory takes on board, is from a disengaged stance. Smiles and scowls were considered as actions, deliberately placed on the face by a calculating subject. Communication was considered as the putting into words of pre-existing thoughts. Action in general was considered on the model of an acting out of pre-existing intentions. In all such cases, the true mindedness of the subject is represented as an inner function, not constituted by the social and praxical activities themselves. But although such disengaged and deliberative action is certainly a phenomenological

possibility, it is not our principle mode of engagement with the world, and such deliberative action is arguably a derivative form, logically parasitic upon the intrinsically mindful expressive everyday potentialities of the agent. However it is just this everyday engaged mode of cognition, our practical reasoning, that is disrupted by schizophrenia. By modelling everyday cognition on disengaged reasoning, cognitivism leaves itself without the resources to explain the deficits in spontaneous mindful engagement with the world characteristic of schizophrenia.

Furthermore, the cognitivist's metaphysics of mind - their thought, affect, and intention - positively invites the failure-of-willed-action account of schizophrenic symptoms. Because the only conception of mind in action that the cognitivist has is that of the disengaged perspective, and because the cognitivist must illicitly presuppose the ground roots of our mindedness, schizophrenic symptoms can only be understood in terms of executive failures, failures in putting an intact, sane, mind into action. The failure of cognitivism to offer a framework for the understanding of the psychotic mind is even more apparent than is its incapacity to theorise the intact mind. This is because the discrepancy of the illicit and *sane* homuncular self within the *psychotic* individual (i.e., the impression the theories give that the subject with schizophrenia is actually fully sane) is more noticeable than its tacit presence within the sane individual.

This chapter has examined the intrusions of cognitivism's metaphysics of disengagement and its alienated third-person epistemology into cognitive psychology. It has argued that the mechanistic and modular preconceptions of cognitivism give cognitive theories the appearance of greater explanatory power and scientific content than they genuinely have. The next chapter investigates a parallel disruptive influence on the psychological theory of psychosis from cognitivism's alienated first-person epistemology.

## Ch. 5. Cognitive Models of Schizophrenic Self-Estrangement

### 1. Introduction

The last chapter examined what might be called the 'ontology' of schizophrenia, which is to say it looked at theories of the *structure* of the schizophrenic mind and of its disintegration as primarily manifest in action and language use. However most cognitive theories of schizophrenia have been principally focussed on the psychotic subject's *experience*, and less on the coherence of their action and thought. In particular, earlier cognitive models of schizophrenic hallucination theorised this symptom as a defect in ordinary perception, as a defect in a cognitive 'filter' allowing irrelevant stimuli into consciousness, a failure in discrimination or a bias in the interpretation of stimuli. The more recent theories investigated in this chapter locate the experiential deficit in what they refer to as the subject's *self-consciousness*, by which is meant not that everyday function of feeling awkward under the gaze of others, but rather the very same 'introspective' faculty of 'inner sense' posited by the cognitivist which is supposedly required for our knowledge of our own minds.

Epistemological rather than ontological issues, therefore, come to the forefront in such theories, and also in this chapter. Here the majority of attention is directed toward the philosophical theory of hallucination and thought insertion provided by George Graham and G. Lynn Stephens as well as to Frith's psychological theory of the same phenomena. What I shall argue is that there is little hope of successfully translating the cognitive theories out of their epistemologically dubious cognitivist idiom, and that the theories only appear to explain the experiential abnormalities of schizophrenia by positing a breakdown within normal subjectivity *that has itself been modelled in a fundamentally alienated way*.

The cognitive theorist's focus on disturbances in experience also - it shall be argued - allows for a *tacit assimilation of psychosis to normal mindedness*, making way for an appearance of explanatory prowess drawing on the paradigms of folk psychology. This is because the cognitive psychologist's conception of experience draws heavily on the cognitivist's construal of perception as *input*, as something prior to and conceptually distinct from the contents of the mind, distinct from the subject's belief, knowledge and understanding. With this conception of experience in place, it is open to the



psychologist to interpret delusion as an attempt by a rational subject to explain their abnormal experience. There is of course nothing wrong with the idea that strange beliefs may arise from abnormal experience, but it is questionable whether the bizarreness of the *form* of schizophrenic delusions can be successfully theorised in terms of the strange *content* of hallucinations, and whether the strangeness of schizophrenic hallucinations can itself be understood solely in terms of their experiential content.

These considerations about the character of delusion will be addressed in chapter 6. Chapter 5 is concerned just with the construal of hallucination and the experiential ground of the delusion of thought insertion as a disorder of 'self-consciousness'. The focus is once again on the influence of mechanistic mentalism (*cognitivism*) on the cognitive psychological theories, but because the concerns are primarily epistemological rather than metaphysical it is the mentalism rather than the mechanism that requires the main of the attention. For its critical stance the interpretation presupposes the arguments against cognitivist accounts of self-ascription given in chapter 3 section 3 (section 2. i below provides a brief rehearsal). Section 2 of chapter 5 predominantly investigates philosophical theories of psychotic experience and argues that their appearance of explanatory potential presupposes an untenable mentalist metaphysics - and that when this is exposed their explanatory power is revealed as minimal. Section 3 investigates cognitive psychological theories and suggests that they suffer from the same defects. A final section uncovers the alienated roots of both the philosophical and psychological theories.

## **2. Philosophical Theories of Psychotic Experience**

### **i. Immunity to Error?**

In chapter 3 it was argued that the immunity to error that we enjoy in sensation and propositional attitude self-ascription is not underwritten by the perspicacity afforded by any faculty of 'inner sense', nor by a perfectly functioning 'self-monitoring mechanism', but rather by *grammar*. This is to say that talk of error or correctness is here simply ruled out by the normative structure of first-person folk-psychological discourse, which in this respect differs from many other language-games employing truth-stating factual propositions.

A similar observation about immunity to error has often been made about the *subject* - the 'owner', as the literature would have it - as well as the content of attitudinal self-ascriptions. The claim is not that we cannot be deluded about who we are - in the sense of knowing our own name, status, age, relationships, etc., but rather that a question such as 'Someone is doubting, but am *I* the one who is doing it?'<sup>128</sup> is not merely unlikely to be asked, but is rather *unintelligible*. Our immunity to error does not involve our always being *right* that it is we who have our experience; rather: the owner of the experience *just is the subject who avows it*. As Ryle obliquely puts it<sup>129</sup>:

'I' is like my own shadow; I can never get away from it, as I can from your own shadow. There is no mystery about this constancy, but I mention it because it seems to endow 'I' with a mystifying uniqueness and adhesiveness.

This permanent shadow appearance of the 'I' is, as Wittgenstein would say, a shadow cast by *grammar*<sup>130</sup>; furthermore the appearance depends upon casting in a material mode what is really a grammatical or conceptual function. Thus we would do better to say, for example, not: 'I cannot be mistaken that it is I who am in pain', but that there is no room for talk of mistake or correctness, error or rightness, in the language game of folk-psychological self-ascription.

Such grammatical facts, however, seem to run up against the phenomenon of *thought insertion*<sup>131</sup>, natural descriptions of which urge on us not only the possibility but also the actuality of error as to the ownership of experience. Consider K. W. M. Fulford's description of thought insertion<sup>132</sup>:

The claim that one is *having* an experience which at the same time is not one's *own* experience seems to be almost self-contradictory. Indeed the 'adhesiveness' of experience makes it, as Ryle said, as 'inescapable as a shadow' (Ryle, 1949). I can imagine myself separated from parts of my body, perhaps even from my body as a

---

<sup>128</sup> Glover, *The Philosophy and Psychology of Personal Identity*, p. 63. C.f. Wittgenstein, *The Blue and Brown Books*, p. 67.

<sup>129</sup> *The Concept of Mind*, p. 198.

<sup>130</sup> And not, that is, a mundane reflection of a super-mundane or metaphysical truth.

<sup>131</sup> And other passivity experiences too.

<sup>132</sup> *Minds and Madness*, p. 6.

whole. But I cannot imagine being separated from my conscious experience. Yet this is exactly what the schizophrenic patient with thought insertion *does* experience.

Now what Ryle was talking about was really the 'adhesiveness' of the 'I' (the pronoun, that is), and not of experience. But this only serves to make the point more strongly. For the schizophrenic with thought insertion is most naturally described not so much as experiencing themselves removed from their experience, but rather as delusionally believing that they are not the owner of (what is in fact) their own experience.

The psychiatric phenomenon of thought insertion thereby *appears* to clash with a Ryleian or Wittgensteinian account of the grammar of 'I', and it seems that one or the other will have to give. Thus Fulford (p. 16) argues that to the extent that philosophers have simply *assumed* the 'adhesiveness' of experience (or, as we might want to rephrase it, of the 'I'), the phenomenon of thought insertion has a 'wide significance' for the philosophy of mind. The philosophers Graham and Stephens (henceforth G&S), who have provided the principle philosophical account of thought insertion and schizophrenic hallucination similarly and reasonably suggest that our understanding of a phenomenon can often be improved by looking at its breaking points, as 'unstressed or orderly psychological activities often conceal their component structures or elements'<sup>133</sup>.

The general tack taken by philosophers and psychologists is to shape their account of the mind in a way which would render the patient's reports of thought insertion coherent. The appearance of contradiction in what the psychotic subject says is removed and understanding is thereby reached, perhaps in much the same way that an apparently paradoxical or contradictory utterance issued by a sane subject can be made coherent by the offering of a further story or more detailed analysis. By providing an account of our self-ascription which makes room for the possibility of error, the reports of subjects with schizophrenia that thoughts occur in their minds which are not their own can be viewed as intelligible mistakes. By making what the subjects say intelligible in this way the theorist aims to make the symptom itself psychologically intelligible.

---

<sup>7</sup> *When Self-Consciousness Breaks*, pp. 2-3.

The attempt to render understandable what the psychotic subject says is of course an admirable one, and any theory which truly enables empathy with and understanding of the schizophrenic condition is to be welcomed. But in opting so swiftly for an account of self-ascription that is so accommodating of the schizophrenic self-ascription there is a danger of rendering the symptom *too* intelligible. That is to say, in being viewed as a sane response to an abnormal experience the symptom may not be so much understood as re-theorised; its particular psychotic quality overlooked, and the schizophrenic condition assimilated to one of normality<sup>134</sup>. The danger in short is one of ignoring what is ultimately psychotic about psychosis. Paradoxically, an attempt to understand psychotic phenomenology which removes the appearance of bizarreness in psychotic self-ascriptions can thereby fail to attain its own goal, for the bizarreness which is removed is in fact constitutive of the condition.

The psychiatric textbooks make what could be considered a start on the cognitive theorist's enterprise of rendering thought insertion coherent. They typically draw a distinction between 'having thoughts in your own head, thoughts consciously present to you, and thoughts which in this sense *you* are thinking, yet which you experience at the same time as the thoughts of some *other person or agency*'.<sup>135</sup> Philosophical accounts of thought insertion have largely been geared to unpacking the *meaning* of this distinction, whilst cognitive psychological accounts tend to take the distinction for granted, and try instead to explain *why* the schizophrenic subject experiences their own thoughts as belonging to some other agency.

The principle philosophical account of thought insertion is provided by Graham and Stephens (henceforth G&S)<sup>136</sup>. G&S aim to remove the appearance of contradiction, in the description of what they aptly describe as a condition of 'introspective alienation', by drawing a distinction between two forms of 'self-consciousness'<sup>137</sup>: one of 'realizing that something occurs in my mind', the other of 'experiencing it

---

<sup>134</sup> In this respect it is perhaps not surprising that cognitive accounts tend to play down the *delusional* character of thought insertion, reclassifying it as a purely experiential anomaly. Cf. Frith p. 80.

<sup>135</sup> Fulford's presentation of the textbook definition. *Minds and Madness*, pp. 6-7.

<sup>136</sup> See G&S: (eds.) *Philosophical Psychopathology*; (authors) *Mind and Mine*; and most recently (authors) *When Self-Consciousness Breaks*.

<sup>137</sup> *Mind and Mine* p. 95.

[i.e. that which occurs in my mind] as agentically mine'<sup>138</sup>. To pre-empt what will follow: What I shall argue is that i) this distinction fails, that ii) it is only made to seem plausible by the tacit employment of a disengaged philosophy of mind and alienated phenomenology of self-ascription (subscribing to the philosophical conception of 'self-consciousness' criticised in chapter 3), that iii) the idea that we can have experiences that are not our own is not (as Fulford writes in the passage quoted above) *almost*, but rather *actually* self-contradictory, and ultimately (especially in chapter 7) that iv) what we need to do to gain a psychological understanding of thought insertion is not to remove this appearance of contradiction, but rather to *come to appreciate the distinctive logical and phenomenological character of the contradiction itself*.

## ii. The Psychology of 'Introspective Alienation' and the Rhetoric of 'Self-Consciousness'

As G&S develop it, the construal of thought insertion as a disorder of introspective alienation involves understanding thought insertion as a failure in the 'normal experience of *introspective awareness*'.<sup>139</sup> According to these authors (pp. 2-3), we have three levels of consciousness, of which the second and third are forms of self-consciousness:

1. The occurrence of thoughts and feelings
2. Our noting of the occurrence of these thoughts and feelings. (p. 7: 'mere introspective awareness of a thought'<sup>140</sup>)
3. Being aware of noted thoughts and feelings as 'agentically' our own. (p. 7: 'my experience of the thought as mine')

Thought insertion, they theorise, reflects a failure in the third level of (self-)consciousness.

---

<sup>138</sup> Ibid. p. 107.

<sup>139</sup> *When Self-Consciousness Breaks*, p. 1. Page references in the immediately following text are to this book.

<sup>140</sup> G&S suppose that all normal self-ascription is made possible by self-awareness, and that this self-awareness is to be understood as a form of introspection. For example, *When Self-Consciousness Breaks*, p. 123: 'Presumably the subject is aware of her inserted thoughts introspectively, just as she is aware of her "normal" thoughts.'

Investigating G&S's theory will involve focussing on this third level, for it is at this level that the problem with thought-insertion is theorised: the subject with thought insertion is aware of thoughts occurring in their mind, but is not aware of them as thoughts that they themselves have produced (and for this or some other reason they hold that the thoughts originate with someone else). But this third level clearly depends on the second – it involves the introspective awareness of thoughts and feelings under a certain aspect, and there are various questions that can be asked about this level.

It is clear from how G&S write that by 'self-consciousness' they do not mean to indicate that empirically indisputable unpleasant feeling of being under the scrutiny of others, and that by 'introspective awareness' they are not referring to, say, an awareness of our character structure gained by soul-searching rather than the interventions of friends. Rather, they seem to suppose the validity of the cognitivist - or more broadly speaking, mentalist - view that our capacity to self-ascribe thoughts and feelings is dependent on a form of internal perception, an epistemic faculty whereby we come into contact with the contents of our minds, contents which, to employ a phrase of Crispin Wright's, are to be understood as 'there anyway', independent of this introspection<sup>141</sup>. This was precisely the conception of (the supposed grounds of) self-ascription that was criticised in chapter 3, where the argument was that this epistemic conception inevitably failed to do conceptual justice to the immunity to error that we enjoy in such ascriptions. Putting the argument briefly, the position there taken was that mental contents are (by and large) *constitutively inalienable*: they *cannot* be theorised in terms of an epistemic perspective that makes talk of error a logical possibility. To pursue the argument in the present context: G&S's manner of making thought insertion intelligible is to view it as a disruption within what is in fact an impossible faculty; this fails to make anything intelligible and fails to deepen our understanding of the psychiatric phenomenology.

This way with the authors is clearly too quick and impatient, for it must be admitted that there is an intuitive plausibility both in a reading of thought insertion as introspective alienation – where by this phrase is meant an alienation within some supposedly everyday introspective capacity<sup>142</sup> (just as there is

---

<sup>141</sup> Crispin Wright, *Wittgenstein's Later Philosophy of Mind*.

<sup>142</sup> Whatever the intuitive plausibility, the argument of this chapter will be that this is the wrong way to understand the genuine psychopathological intuitions behind talk of introspective alienation. Compare in this respect Louis Sass's argument (in *Paradoxes*

in the idea that the ‘splits’ in schizophrenia are between psychological faculties such as thought and language or affect and action) – and also in the idea that we may have to adjust our conception of the nature of the mind because of those concerns that a consideration of the psychopathology forces upon us. Nevertheless, I shall suggest that a consideration of G&S’s rhetoric and argument reveals that their conception of the nature of mind is implausibly alienated and that it is actually only the setting that is provided by what is *presupposed* rather than *argued* concerning the nature of self-ascription that makes their description of the psychopathological situation at all plausible

Consider G&S’s description of the following situation: I am reading a book, but then I stop reading and start thinking how boring the book is; I feel ‘disappointed, maybe even cheated’<sup>143</sup>:

Suppose that you are reading this book, attending to every word, and suddenly you shift attention from the object of your visual experience (the book) to *your* experience of reading. Suddenly, suppose, it seems to you that you dislike the book... This is an example of being aware of your own feelings as your own. You are conscious of yourself as reading, as feeling disappointed, and as cheated. You are, in [William] James’ sense, self-conscious.

Now although this is put forward as an everyday example, it could be urged that it isn’t really a description of anything. For in the normal run of things I do not shift attention from the book to my *experience* - I just start to feel bored by or cross about the book, start to feel that I’m wasting my time. My experience is not an object of my attention, and just because my attention wanders from the book we are not forced to find an alternative object for it. (We are not always equally attentive.) Furthermore it is unnatural and alienated to think that it suddenly ‘seems to you that you dislike the book’, for what is the ‘seems’ supposed to be qualifying here? What it suddenly seems to us is that *the book* is boring, no good, and this ‘seeming’ is *constitutive* of our dislike, and not a mode of attention directed towards a negative feeling. The rhetoric of the example encourages the picture that our thoughts and feelings are objects for

---

of *Delusion and Madness and Modernism*) that the introspective alienation from which schizophrenics suffer is not a defect within a perfectly normal form of self-consciousness, but rather a *pathological tendency to adopt just this mode of introspective awareness itself*. Sass draws this contrast between his own and G&S’s position in his review essay *Analyzing and Deconstructing Psychopathology*.

<sup>143</sup>When *Self-Consciousness Breaks*, p. 2.

our attention, and we are thereby put in the alienated position of requiring a faculty of self-consciousness to find out what it is we think. The same picture is encouraged by the above-quoted remark concerning that function of supposed self-consciousness making possible our 'realizing that something occurs in our mind'. The entifying or reifying rhetoric of mental contents - thoughts as *things*; the spatial conception of mind - as a place *within* which such *things* occur; the passive and objectifying rhetoric for describing our thinking - as a mere *happening* or *occurrence*: such can hardly help but encourage the view that some quasi-perceptual access is required to these inner objects. So we are left 'realizing' that we are thinking, or as we are now forced to put it, that thinking is 'going on in' our minds, as if some genuine cognitive achievement employing some genuine cognitive faculty had thereby been had.<sup>144</sup>

A similarly alienated perspective is provided by Frith in his 'metarepresentational' account of schizophrenia. This account is supposed to unify his failure of action and failure of self-monitoring theories of schizophrenia. I shall not consider it in detail because it has (I believe) since been abandoned. The rhetoric however is very revealing: according to Frith (pp. 115-116) metarepresentation is a 'general mechanism ... that is fundamental to conscious experience':

The ability to reflect upon how we represent the world and our thoughts is the most striking feature of our conscious experience. While thinking what to write at this point, I have been staring straight out of the window. In front of me are many trees. When I become conscious of this activity, what I become conscious of is "me looking at trees". This is the critical feature of conscious awareness. It is not representing "a tree", because I was looking at the tree for some time without being aware of it. It is representing "me looking at a tree". This is representation of a representation and, hence, metarepresentation. I propose that meta-representation is the crucial mechanism that underlies this self-awareness. Self-awareness cannot occur without metarepresentation. It follows that people who have difficulty with metarepresentation must also have an abnormal state of self-awareness.

First it can be questioned whether our capacity to reflect on how we represent the world is the most striking feature of our conscious experience. It can even be questioned whether it *is* a feature of our conscious experience. Surely what is most important is just our capacity to look at/see/be conscious of trees, and not our (rather pointless?) reflecting to ourselves that this is indeed what we are doing. Far

---

<sup>144</sup> There is of course the possibility that we reflect to ourselves that we are thinking, but such reflection is hardly a cognitive achievement - we do not thereby become appraised of anything that we did not already know.



from this being ubiquitous, or 'the critical feature of conscious experience', what matters in our everyday negotiation of the world is that we *see* trees and act accordingly.

Consider another significant feature of this example: Frith is moved to his thoughts on a supposedly reflexive consciousness whilst in a disengaged state of mind, 'whilst *staring straight out* of the window'. When in such a state of mind, when we do start staring and momentarily disengage from our actual perceptual engagement with our environment, the tendency to suppose that we are directly confronted not with an objective world but rather with a 'visual scene', a set of static 'inner images', is all too great. In so doing we tend to reify our perception itself, turning it as it were into an object for another 'meta' form of ('internal') perception. Far from being exemplary of the normal condition, the disposition to stare and for consciousness to become hyper-reflexive are in fact aspects of schizophrenic experience, known in the phenomenological literature as the 'truth-taking stare'.<sup>145</sup> Everyday consciousness, however, knows nothing or very little of such an abstracted and distanced perspective, the constant taking of which disposes *toward* madness rather than wards it off.

There is furthermore the question as to what the true cognitive content of the 'meta-representational' awareness mentioned actually is. What, for example, do we learn that we did not already know when we reflect to ourselves that we are looking at a tree out of the window? We were not unaware before that we were looking at a tree. (What else did we take ourselves to be doing?) If asked later what we had seen out of the window, we should have been able to report on the trees without difficulty. In short, it is hard to see what rôle such a metarepresentational mechanism is actually required to play. There is of course the well-known fact that we can be affected by our environments in ways which depend on our sense organs even though we may remain unaware of the stimuli in question. This however does not mean that when we are perceptually aware of our environment we are simply directing

---

<sup>145</sup> C.f. Wittgenstein's remarks on staring as a psychological prerequisite for much metaphysical speculation on the nature of experience and the self (*PI* para 412, *BB* 66, 176, *NFL* 309). And also Michael Williams' description (*The Good and the True* p. 125) of the genesis of the conception of the disengaged self - of 'the world over there, you over here' - that is important in Descartes' (amongst others) philosophy: 'The picture [of the disengaged self] can seem absurd, but it can be compelling enough when you sit late in your office, all quiet, sniffing your pen; or like Descartes, in your dressing-gown, by the fire, at night, alone'. For the phenomenological perspective on schizophrenia as a disorder of 'hyperreflexivity' - as a proneness to employ, and not a failure within, reflexive consciousness, see Louis Sass's *Madness and Modernism* and *The Paradoxes of Delusion*.

a conscious gaze on otherwise unconscious experiences; rather we direct our conscious attention on the *environment itself*.<sup>146</sup>

To summarise: both G&S's and Frith's theorisation of introspective alienation takes place within the context of a general theory of 'self-consciousness' which depends on a both phenomenologically and epistemologically alienated conception of our everyday experience. It is only by presupposing this suspect perspective however that the logical space is opened up for a theory of introspective alienation as a failure within introspection.

### iii. Thought and Agency: The Action Analogy

Invoking a faculty of self-consciousness does not in itself provide the required conceptual apparatus for understanding thought insertion. G&S propose further that 'the key to rendering introspective alienation intelligible lies in distinguishing the claim that a person is the subject in whom a given mental episode *m* occurs and the claim that a person is the agent or author of *m*.' The patient with thought insertion, then, is to be understood as claiming that whilst a thought is occurring 'in' them and is in this sense theirs, it is not one of their own in as much as they were not its 'author'.

The claim is not that the patient is admitting that the thought they had is not *original* - in this sense we all admit to having thoughts that are not our own - yet we do not normally complain of thought insertion. The meaning of the distinction (between authorship and ownership) is rather given by means of an analogy with bodily activity and action, drawing on Wittgenstein's discussion of the difference between arm raising and arm rising.<sup>147</sup> *Arm raising* is something that I *intentionally* do, but my arm's *rising* is something that may happen whether or not I intentionally raise my arm (someone else may grab hold of it and lift it, for example). G&S explicate this conceptual distinction in experiential terms: '*I have the sense that [my italics] I am actively involved in some of the movements of my body but passive with*

---

<sup>146</sup> The phenomenon of 'blindsight' may seem to demand some kind of explanation in terms of a metarepresentational mechanism.

The temptation can however be resisted. See chapter 7 section 2.1 and John Hyman's *Visual Experience and Blindsight*.

<sup>147</sup> *PI* §§612ff.

respect to others', and they then use this experiential reading to pursue the analogy with thought insertion (pp. 98-99):

Just as I may experience myself as either agent or patient with respect to a particular bodily movement, I may experience myself as actively or merely passively involved in the "movements" of my mind .... This sense of activity or passivity is especially marked in the case of episodes of thinking. Thinking a certain thought or thoughts is something that I often feel I do voluntarily, even deliberately. I may decide to recite slowly the last line of Swinburne's "In the Garden of Proserpine"; I may cogitate with the intention of rehearsing Anselm's ontological proof for the existence of God. More commonly, I have a sense of deliberately directing my thinking toward a certain project or theme, such as crafting an apology, finding a solution to a problem, recollecting my trip to Berlin or Tenerife, without having some specific sequence of thoughts in mind at the outset. On the other hand, I may feel that certain thoughts occur in me through no doing of my own. The lyrics to an odious advertising jingle may run through my head unbidden and may continue despite my efforts to dismiss them. The name "Rosebud" may pop up in my stream of consciousness without my being able to see how its occurrence is relevant to any cognitive project in which I am currently engaged. .... In our view, admitting that thought *m* occurs in my mind while denying that I think *m* is like acknowledging that my arm went up but denying that I raised my arm. I accept that I am the subject in whom *m* occurred but deny that I am the agent responsible for *m*'s occurrence.

The analogy may appear straightforward, and it's promise is one upon which G&S base the whole of their case for an understanding of thought insertion as an alienation within a normal faculty of introspection, but several questions can be raised as to its validity. Consider first that which is given as the analogical ground: our experience of ourselves as either agent or patient of bodily movement. To be sure, there is a *logical* distinction to be had between, in Wittgenstein's example, arm raising and arm rising, but a reflection on our everyday action doesn't reveal that this distinction becomes manifest in our *experience*. We are shocked when our bodies move when we don't expect them to, and this could be considered an experience of passivity, but as concerns our everyday pre-reflective action there seems to be *very little* by way of an actual experience of agency. What we rather have, it might be argued, is an *absence* of any unusual experience, and not necessarily – as some phenomenologists have argued<sup>148</sup> – the

---

<sup>148</sup> Josef Parnas, for example, in *The Self and Intentionality in the Pre-Psychotic Stages of Schizophrenia: A Phenomenological Study*, suggests that our basic pre-reflective agency is constantly experienced, albeit not as an *object* of awareness for the self, but rather as the self's non-relational self-awareness. Parnas considers this (p. 6) a 'basic dimension of subjecthood, technically called

*presence* of a particular experience of agency. To successfully argue that we nevertheless *do* have such an experience of agency operative for what would be almost all of the day would require the provision of grounds for asserting the existence of such an experience. On the one hand such grounds ought not to simply involve an appeal to supposed *disruptions* of this experience, for these could be more parsimoniously explained simply in terms of the *abnormal presence* of indisputable experiences of failures in action or interferences with our bodies.<sup>149</sup> On the other hand they cannot simply invoke sub-personal facts (such as the re-afference copy in the visual system discussed in chapter 2), for what is wanted is warrant for describing any capacity of the body to react differently to (for example) internally initiated movements than to environmentally induced movements in psychological and not simply physiological terms. The question of what such grounds might be remains open.

Nevertheless, the experiential rhetoric may perhaps be dispensable; there is after all certainly a *logical* distinction to be had between action and movement. The question to be considered now is whether this distinction can properly be drawn within the mind – where we are concerned with thoughts and not bodily actions. In giving a positive answer to this question G&S appear to rely on some opaque turns of phrase and thought. They note, providing some rather strange examples (*cogitating with the intention of rehearsing* Anselm's ontological proof...), that we pursue goals in thought just as in action. I intend to think something through – a mathematical proof, a poem, a philosophical argument, a political debate – and I do. But such examples have as their opposites cases in which I do not *beforehand* form the intention to think what I do, and so they don't provide the necessary contrast between agency and passivity in thought. To draw a comparable contrast with action it is necessary to distinguish between

---

*ipseity*'. In what follows it should become clear that whilst I sympathise with and approve of the project of avoiding objectifying conceptions of experience or thought which alienate the true self from these by making them objects of experience for the self (as the cognitivist accounts do), I do not share the view that there is even a pre-reflective or tacit experiencing of the self by itself – that there is a characteristic experiential structure to everyday lived embodiment. Schizophrenic disruptions are not to be thought of – in my view – as disruptions to *any* form of self-consciousness.

<sup>149</sup> A silly example by way of comparison: saying that I may sometimes experience a cow in front of me (when, say, there is a cow in front of me) does not entail that I normally have an experience of not having a cow in front of me – in such cases I have no experience at all, or have some other experience.

thoughts that are akin to intentional actions *whether or not* they are premeditated and thoughts that are akin to mere movements.<sup>150</sup>

The contrast that G&S employ, then, is between deliberately directed thought – whether or not the direction was stipulated in advance – and thoughts that ‘occur in me through no doing of my own.’ In the first category might come the spontaneous crafting of an apology, and in the latter, random catch-phrases or advertising ditties. It is whatever contrast such examples provide that gives the content to G&S’s distinction between agency and passivity in thought.

A first worry about this distinction is that it is forced on the reader more by the unnatural rhetoric employed than by its intrinsic plausibility. The concern, that is, is that it is the construal of thoughts as ‘episodes’ or mere ‘occurrences’ that take place ‘in’ us<sup>151</sup> that *seduces*, rather than *rationally persuades*, the reader into a view of thinking as *in itself passive* and as requiring supplementation by incorporation within *voluntary*, perhaps even *deliberate*, cognitive *projects* or imaginative *pursuits* before it can wear its familiar homely face. In a more natural idiom, for example, thoughts do not occur *in* us but *to* us. Furthermore, it is doubtful whether we would normally call an occurrence such as the name “Rosebud” irrelevantly coming to mind the having of a *thought*, and the appellation may even seem a bit stretched for longer snatches of phrases. In so far as such mental occurrences are, as we characteristically say, *mindless*, they could be said to fail to be thoughts for just this reason. (We can imagine, in reply to a question: ‘What are you thinking?’, the answer: ‘Nothing, but I’ve got an irritating advertising ditty on the brain.’)

---

<sup>150</sup> There would be several risks involved in drawing the action analogy in terms of premeditated or unpremeditated thought. One would be the requirement that the specification of the intention would need to make reference to the thought itself – in which case the thought is already contained in the intention and does not need thinking. Another would be the requirement that forming an intention to have a thought would itself require prefacing with another intention – to form this intention, leading to a regress of prior intentions. The examples that G&S give show that it is only on rare occasions that we form prior intentions to have thoughts; a comparison with characteristic cases of thought-insertion reveals that the occasions are not the same.

<sup>151</sup> Consider too the frequent talk, in American philosophy of mind and action, of actions as *events* – instead of as, say, *deeds*. For, pursuing a natural enough analysis (c.f. Michael Morris’ *The Good and the True* p. 136), events are things that merely *happen* (happen *to* us, for example), whereas deeds are *done* (done *by* us). Hardly a refutation of an entire field of action theory, admittedly (!), but enough to generate suspicion as to the metaphysical credentials of what is often spelt out merely *using* (– without *interrogating*) the rhetoric.

Any straightforward analogy with bodily action will fail for the following reason: Actions involve bodily movements, movements which can also occur independently of the actions of which they are at times constitutive. As with movements, mental imagery (phrases running through the mind etc.) can also occur both with and without thought. But much thought occurs without any such mental imagery – we can both think and come to realise many things without mental imagery of any sort, and sometimes the mental imagery can be quite incidental. Of course we can also act without moving: we may be hiding or standing on one leg quite still. For most actions however the bodily movements are not mere accompaniments but rather essential components; this is not the case with thoughts.

Two clarificatory options could be taken at this point. Firstly, it could be suggested that the relevant contrast is between mental imagery that expresses thoughts and mental imagery that does not. This however doesn't provide the distinction between activity and passivity in the mind that is stressed by G&S. Their distinction was between thoughts that we direct in a certain way, and 'thoughts' that pop into our mind unaided. (This contrast makes it hard to know how to characterise much of our thought, such as our undirected ruminations, that are nevertheless not out of the blue.) It also makes it hard to understand the idea that our agency is expressed (in thought) in our deliberate shaping of our thoughts, in our *crafting* of apologies, in our problem-solving or recollections. (We shape or direct our thoughts, not our mental imagery.) Furthermore it creates the additional theoretical requirement of providing an explanation as to why the subject with thought-insertion believes that their mental imagery which is experienced as being out of the blue is nevertheless expressive of thoughts (G&S do in fact provide such an explanation, which I shall consider in the next section). This of course takes for granted that all cases of thought-insertion are cases where the thought that the subject is claiming to have been inserted was expressed in *imagery in foro interno*. (It may be that some 'inserted thoughts' are not accompanied by mental imagery, although this is *perhaps* unlikely.) Finally, whilst this option paves the way for a psychological understanding of thought insertion, it seems to distance the phenomenon from other similar psychotic symptoms which will now require a different form of explanation. I am thinking of the other experiences of introspective alienation, of having not merely the thoughts but also the *feelings* or *urges* of others; in these cases there does not appear to be any neutral substrate (equivalent to the verbal imagery) which could occur with or without the emotion.

The second clarificatory option would have it that it is genuine thoughts and not mere mental imagery that is at issue. This enables us to stick more closely to the clinical details as they are commonly presented, and also allows the contrast between thoughts that pop into our head and thoughts that we consciously direct to be maintained. It is however unclear that any such passive/active contrast can be maintained of the majority of our thought which is neither particularly directed nor really unbidden. It is also unclear that even in such cases thoughts can truly be considered actions. To be sure thinking shares the surface grammar of sailing or writing: it is something that we *do* and is in *this* sense an action. But then again *hoping* and *wishing* are things that we do, and it is hard to see how an active/passive contrast could be sustained in these contexts. Similarly with thought: there is a sense in which thought 'just happens' or unfolds of its own accord, where to think of thought *either* as something that we 'do' *or* as something which we 'witness' is to place the thinker *behind* the thoughts and not in their midst.

The intuition behind this repudiation of what a Buddhist may refer to as the illusion of the 'ego' can be spelled out analytically. When we consider (normal - i.e. bodily) action, the intentionality of the action is manifest in its accordance with the agent's intention. If I ask you what you are doing you can tell me - you tell me the intention you have in acting which intention defines the action as the action it is. The aim of the action is thereby given. But when we consider intention or thought itself there is nothing else other than the thought itself - no further thought or intention is needed to specify the content of the thought, to specify the thought as the thought it is.

Consider a typical thought: it occurs to me, as I'm lying in bed, that I really must get up now. Is this thought active or passive, intended or unintended, directed or undirected? Whilst the movements made when I get out of bed may or may not be components of actions (perhaps I am pulled out of bed, or perhaps I get up of my own volition), the thought that I ought to get up is not something I intended to 'do'. It was not intended but yet also not unintended; it did not occur out of the blue (it was not unexpected) but it was also not planned (or expected). Such an active/passive contrast does not it seems apply to the mind itself, except in those rare instances (such as those examples provided in the quote from G&S) where we may plan to have thoughts. Again, the argument here does not trade on illicitly taking as the ground of the action model only those actions which are actually premeditated, intended beforehand. The worry isn't that there will be an infinite regress of *prior* intentions if thought is

construed as an action; the threat is rather of a regress of *concurrent* intentions against which the activity or passivity of the thought could be measured.

To conclude the argument: if the focus of G&S's theory of thoughts being experienced as intended or unintended is concerned with genuine thoughts *per se* and not just inner imagery, then the active/passive (intended/unintended) distinction cannot be sustained. The analogy with arm raising and arm rising which underlies the treatment of thought as an action breaks down. Thought may be described in idioms that suggest activity or passivity (thoughts (passively) occur *to* us, or (actively) we *think* of this or that), but such different surface-grammatical modes fail to reveal any relevant depth-grammatical (conceptual) commonality with actions. There will, then, be no possibility of someone suffering from schizophrenia experiencing a thought as unintended, for there is no such thing, when considering everyday thoughts, as intending or not intending what we think.

#### **iv. Thought and Ownership: The Attribution of Thoughts**

The above considerations rejected the view that the concepts of activity and passivity, or experiences of oneself as either agent or patient, can be considered conceptual currency when we are considering everyday thought rather than action or movement. Nevertheless two coherent proposals were identified: sometimes we have phrases rattling through our minds (which may or may not express thoughts), and sometimes our thoughts are more out-of-the-blue than at other times. The question remains as to why the subject suffering thought insertion views these more-or-less random snatches of inner speech as expressing thoughts, or why out-of-the-blue thoughts are liable to be attributed to another agency.

Following a popular American approach, G&S suggest<sup>152</sup> that 'our sense of ourselves as persons and agents engaged in a particular life depends on our proclivity for constructing self-referential narratives, or hypothetical explanations that organize disparate events into coherent projectible patterns'. As they say, 'consider Sam':

---

<sup>152</sup>*Mind and Mine*, p. 101.



Sam is walking in a certain direction because he wants to get some laudanum, believes the best way to obtain laudanum is to buy it at the pharmacy, and believes that the store lies in the direction in which he is walking. He anticipates that upon entering the pharmacy, he will take out his wallet, because he believes that laudanum costs money and that his money is in his wallet, and he intends to purchase the laudanum with the money in his wallet. Sam's behaviour seems sensible and predictable to him because he takes it to be an expression of his own underlying beliefs and actions. He therefore regards it as something he does, as his action, and sees himself as an agent.

The idea, then, is that we normally self-attribute our behaviour, and view it as *our action*, because it seems to us to express our intentional states. If this is right, then we may have an explanation as to why our movements may at other times be other-attributed: Sam may occasionally find himself unable to account for his 'ambulatory behaviour', find himself without beliefs that explain his action, and so and in such a case the behaviour will not appear to him to be his own action.

Once again the account of action is used as an analogy for thought.<sup>153</sup>

[W]hether the subject regards an episode of thinking occurring in her psychological history as something she does, as her mental action, depends on whether she finds its occurrence explicable in terms of her theory or story of her own underlying intentional states. For thinking, as for overt behaviour, having a sense that you are *doing* something involves a sense of *what* you are doing and *why* you are doing it. I find occurring in me the thought that a good dose of laudanum would really hit the spot right now. Do I regard this episode as my action, something that I think, or do I dismiss it as mere verbal imagery running willy-nilly through my head? The answer depends upon whether I take myself to have beliefs and desires of the sort that would rationalise its occurrence in me. If my theory of myself ascribes to me the relevant intentional states, I unproblematically regard this episode as my action. If not, then I must either revise my picture of my intentional states or refuse to acknowledge the episode as my doing.

Whether or not we are to consider what are construed here as thoughts as truly thoughts or instead as mental imagery, the idea is clear: we regard our thoughts as our own actions (or regard our mental imagery as expressive of our thoughts) if they are rationally intelligible in terms of our beliefs or theory about our intentional states.

---

<sup>153</sup> Ibid. p. 102.

And why are the thoughts which a schizophrenic considers to be inserted into their minds not merely disavowed but externally attributed?<sup>154</sup>

A possible explanation for such an extraordinary hypothesis would be that, despite her conviction that an episode of thinking does not express her own intentional states, the episode might seem to her quite intentional. It may be topically relevant - for instance, speaking to concerns that she acknowledges in herself. Unlike nonvoluntary verbal imaginings, such as snatches as doggerel running unbidden through her head (which are notable for their lack of connection with, and tendency to distract from, one's current concerns), the contents of alienated thoughts tend to be personally salient for the subject. ... They mean something to the person, they tend not to be isolated from a person's interests or concerns. Likewise, in the case of voices at any rate, alien thoughts typically exhibit the sorts of grammatical forms appropriate for conversational or communicative speech. They are frequently in the second person, for example. ... Often they are in the imperative mood. ... Their content is appropriate to communicative acts like giving advice or criticism, issuing threats and orders, offering condolence or encouragement.

To summarise: the patient with thought insertion finds themselves having thoughts which do not seem to express their intentional states, but which nevertheless seem to possess the kind of coherence and personal relevance characteristic of thoughts, and for this reason attributes them to another agency.

I shall comment on the logical structure of this form of explanation in section 4, and restrict discussion for the present to the phenomenology. For Sam's behaviour, it must be said, is phenomenologically suspicious; in particular it seems indicative of a considerably *alienated* consciousness. Sam seems to treat himself as an object, or at least as another person. First he 'anticipates' that he will take out his wallet on entering the pharmacy. This however is not normal behaviour. Being free to do as we choose, we are in no need of *anticipating* our own actions. Sam will not be surprised if he doesn't get out his wallet, for if he doesn't then it will be for a reason. He may anticipate the reactions of *the pharmacist* to his request, but to anticipate his own would be to treat himself as someone else.

Second, Sam appears to be in need of a reason for finding his own behaviour sensible and predictable: he finds it predictable and sensible because it seems to cohere with his idea of who he is. Once again Sam has become an object in his own mind. But why should Sam - of all people - need a

---

<sup>154</sup> Ibid. p. 105.

reason for finding his *own* behaviour sensible, and why should he need to *predict* his own actions? Presumably this is because he does not *know* that his action is sensible, or know what he is going to do. But why should he not know these things? With respect to the sensibleness of the action, Sam cannot be unaware of the spirit in which he undertakes it (since the 'awareness' here reflects not an epistemic facility of Sam's but rather the fact that his sincere say-so is stipulative of the character of the undertaking). With respect to knowledge of what he will do, either Sam has made up his mind or he hasn't. But in either case we do not have to do with prediction: if his mind is made up already then prediction will be superfluous, and if he does not yet know what he is going to do then what is required is not a prediction but a decision.

Similar considerations become relevant when the topic of self-attribution is considered (the second quote). If I were asked why I regard one of my 'episodes of thinking' as something I 'do', I should hardly reply that it is because I find its occurrence explicable in terms of some 'theory' I have about my 'underlying states'. Even if we restrict our attention to indisputable action, it seems more plausible to suppose that I count certain behavioural episodes as my actions for *no* reason, than because of such and such. If one of my behaviours surprises me (*and* not just because it constitutes success in a difficult endeavour – hitting a six in cricket, for example, or hitting the bull's eye in darts) then I will not see it as an action; but this is not to say that I should be expected to give reasons why I am normally *not* surprised by my behaviour<sup>155</sup>. Supposing that we are required to 'take' our behaviour a certain way suggests that we are starting from a standpoint in which such behaviour is not inalienably *our own*, or that we need to inspect our bodily behaviour to find out what it is that we are doing at any one time.

Consider too: 'I find occurring in me the thought that a good dose of laudanum would really hit the spot right now.' Really? In the normal run of things, far from introspecting and discovering, I simply *think* that a good dose of laudanum would... Grounds for self-attribution are not required, because the fact of our actually *thinking* the thought removes any need for introspection and the employment of theories of agency. The description too of the person with schizophrenia's claim (that thoughts are being inserted into their minds) as an 'extraordinary *hypothesis*' also misrepresents the clinical phenomenology. The schizophrenic does not offer their claim as an hypothesis, as the product of a speculative piece of

---

<sup>155</sup> C.f. *PI* §628.

reasoning about what might be happening within their minds, but rather believes in it with the intensity of a *delusional conviction*.

The final component of the theory is perhaps the most successful, and could be empirically tested without too much difficulty. All that needs to be investigated is whether thoughts which distract a psychotic subject are more likely to be attributed to external agencies if they are indeed coherent and well-structured. This is not perhaps to address the question as it was put, directed as it is toward a statistical generalisation rather than a rationalising explanation. Taking the issue in this way also precludes the further question as to whether coherent thoughts are more likely to be externally attributed *because* of their coherence. But then it is difficult to imagine how this question could possibly be answered, except by asking for the opinion of the patient. And when the patient is unable to reply to the question, or gives a purely delusional response, it is hard to see any force in the question itself, given that its sense is specified by the kind of answer it receives.

The general perspective G&S employ to understand thought insertion, as outlined in the above three subsections, is persistently alienated. Although it is natural enough to approach phenomena such as thought insertion as the product of a breakdown in psychological faculty, the requirement must first be met of finding and accurately describing such a faculty. G&S provide a general theory of the 'introspective awareness of thought' which only describes an *alienated* consciousness, one in which thoughts become objects for their thinkers. Secondly they develop an analogy between thought and action which once again inserts a phenomenological and logical gap between the thought and the thinker. Thirdly this is backed up by a conception of our relation to our actions which renders them akin to the actions of others. At no point is the alienation implicit in the general perspective obviated by an attention to particular empirical details which would allow the theory to transcend its own theoretical limitations. It now remains to be seen whether Frith's psychological 'output' theory of the hallucination of voices and thought insertion can escape the limitations imposed on it by the epistemological framework criticised (in subsection ii) above.

### 3. Cognitive Psychological Theories of Psychotic Experience

In *The Cognitive Neuropsychology of Schizophrenia* Christopher Frith develops an account of psychotic hallucination as a defect of inner sense (this is his 'output theory'), an account which generalises to abnormalities in thinking such as thought insertion, withdrawal, blocking, broadcast etc. (this is his 'metarepresentational theory'). Just as G&S's work provides the principle philosophical account of thought insertion, Frith's is arguably the most well-known of the psychological accounts of psychotic experience<sup>156</sup>, and for this reason attention will be restricted to his theory. As with G&S's theory, I shall argue that the explanatory structure is vitiated by a commitment to the alienated conception of self-ascription embodied in the very idea of inner sense or inner experience. But because Frith's work is psychological rather than philosophical, I shall read it with a greater application of the interpretative principle of charity, to some degree reading the theoretical content 'back out of' the experimental details rather than vice versa. Nevertheless, it shall be argued that the cognitivist rhetoric and framework cause considerable mischief for the psychological theory. When the interpretative principle is applied, the theory we are left with – or so it is argued – becomes disunified and largely neurological rather than psychological in character. The argument will be that it is only the epistemological framework provided by cognitivism – that framework which I have suggested to be philosophically suspect in as much as it provides a fundamentally alienated conception of normal subjectivity – that lends the appearance of psychological pertinence and theoretical unity to the empirical material.

#### i. Psychological Theories of Hallucination

Frith's 'output' theory of hallucination is developed by means of a contrast with 'input' theories (pp. 68-71).<sup>157</sup> Input theories suggest that schizophrenic hallucinations are due to the *misperception* of ordinary sounds, voices and the like. Perhaps, for example, subjects suffering from schizophrenia are

---

<sup>156</sup> Though see also Ralph Hoffman's *Verbal Hallucinations and Language Production Processes in Schizophrenia*.

<sup>157</sup> All page references in what follows are to *CN*.

worse at discriminating between noises and words, or words and other words, in a similar way to that in which a non-psychotic subject may find it hard to follow a conversation in a noisy bar. Or perhaps the hallucinating subject has abnormal perceptual *bias* - where by bias is meant, for example, the tendency we have to suppose our name called when it hasn't been, or the tendency (p. 71) a mother may have to misperceive sounds as her baby's cries. In short, input theories conceive of hallucination as more akin to illusion or misperception than to what is normally thought of as hallucination. It is not so much that the schizophrenic sees or hears something that isn't there, than that they misperceive something that is there.

Input theories are contradicted by the paucity of sustained evidence for abnormal bias in schizophrenics, and the fact that schizophrenic hallucination can occur in completely quiet surroundings (pp. 70-71).

Output theories of schizophrenic hallucination suggest, by contrast, (p. 71) 'that the patient is talking to himself, but perceives the voices as coming from somewhere else'.

The theoretical success of an output theory of schizophrenic hallucination depends on its being more than the negation of input theories. In other words, 'talking to oneself without realising it' cannot be a simple *gloss* of what it means to speak of 'hearing voices' (and a denial that hearing voices is a matter of misperception rather than hallucination) - it must be a theory of *why or how* schizophrenic hallucination occurs. One might think that talking to oneself without knowing that we are doing so is simply an explanation of the meaning of 'hallucinate', but the output theory presented by Frith seems to have more content than this. Evidence for the theory comes in part from studies of the correlation of the tonal quality and content of the involuntary whispered speech of subjects with schizophrenia with the quality and content of their 'voices' (hallucinations). Such a correlation has, in a few studies, been found to be positive (pp. 71-2).

This finding is not however intended to exhaust the scope of the theory; Frith suggests (p. 72) that it 'is, of course, possible for subvocal speech to occur in the absence of any detectable sound or muscle activity [in the throat, lips, larynx etc.]'. Such subvocal or inner speech is what occurs when we repeat a 'phone number over and over to ourselves in our head in order to remember it. With respect to schizophrenic hallucinations, Frith's suggestion is that psychotic subjects are failing to 'monitor' the origins of their own inner speech (p. 73):

If hallucinations are caused by inner speech, then the problem is not that inner speech is occurring, but that patients must be failing to recognise that this activity is self-initiated. The patients misattribute self-generated actions to an external agent. I have called this a defect of "self-monitoring" (Frith 1987) because the patients are failing to monitor their own actions.

By way of indicating the importance of self-monitoring systems for perception, Frith discusses the mechanism of corollary discharge important in visual perception. This was discussed above in chapter 2. To recap: whenever we move our eyes, the 'image' on the retina changes, as it also does when an object moves past us. Yet only in the first instance is the object perceived as moving. This is because a 'corollary discharge' signalling our 'intentions' is sent to a 'monitor system' when a 'message' is sent to the eye muscles. On the basis of the corollary discharge, motion of the 'image' is 'expected', compensation occurs, and it is 'perceived' as stationary. Disruption to this system can occur when changes to the position of the retinal image which are due neither to the movement of objects nor to intentions to change where we look obtain; curare, for example, can paralyse the eye muscles, and can lead to the impression of movement of perceived objects even when neither the eyes nor the objects actually move.

The psychological rhetoric ('image', 'message', 'perceived', 'expected' etc.) by which the mechanism of corollary discharge is described has already been criticised in chapter 2, but the physiological details which underlie the possibility of the visual perception of movement are unarguable. Whether these physiological details can provide the framework for a psychological account of auditory hallucination remains to be seen.

## **ii. Critique of 'Output' Theories of Hallucination**

When we remain at the level of vocal or subvocal, rather than purely inner, speech, and at the physiological rather than the psychological level, there are doubtless feedback mechanisms to be found that, for example, make for differential responses to the sounds of our own voices and the sounds of others (p. 91). If the output theory is defined in such terms it may or may not be a success: the content of

voices that are heard is clearly identical to what a hallucinating subject may subvocalise on occasion. It is however doubtful that such subvocalisation *always* accompanies auditory hallucination. Furthermore, casting the output theory in this manner would lead not to a cognitive *explanation* of thought insertion but rather to the suggestion that as a matter of fact schizophrenic hallucination is always *accompanied* by movements of the speech musculature.

Frith's theory is not however and in any case restricted to overt speech. The claim is that the subject who hears voices characteristically mistakes their own *inner* speech for the speech of someone else.<sup>158</sup> But the difficulty now is to stop this claim from collapsing into a definition of hallucination. After all, it might rhetorically be asked, what is it to suffer an auditory hallucination other than to think that someone has talked to one but in fact to have 'talked to oneself'? If no-one else has said what we thought we heard, then we must have said it ourself.

One way of trying to save the theory would be to attempt to draw some kind of distinction between talking to and listening to ourselves *in foro interno*. If this were possible then it would be possible to theorise schizophrenic voices as the product of a subject listening to their own inner speech without realising that it was in fact something they had said to themselves. The listening could be a passive component to any normal episode of inner soliloquy, and the talking the active generative aspect. This however would be confronted by the objections raised in chapter 3 and in section 2 above; it would presuppose that we require a form of 'epistemic access' to our own inner musings, that our subjectivity does not have constitutive inalienability as its key form, and that the apparent grammatically sanctioned authority of first-person ascriptions is illusory.

Frith's presentation of the theory is however (pp. 72-3) set in the context of Alan Baddeley's theory of short term memory which employs the notions of the 'inner voice' and the 'inner ear'. If this distinction turns out to be empirically rather than metaphysically motivated it is perhaps possible that it may 'save the appearances' for the output theory.

---

<sup>158</sup> This is to simplify somewhat; 'voices' that are heard by an hallucinating subject with schizophrenia are often not experienced as coming from particular bodily people in the actual physical environment, and they need not be experienced as having particular spatial locations (contrast Korsakoff's syndrome hallucinations).



Unfortunately the details of the theory and the issue of their correct interpretation are not clear, but in brief: Baddeley notes that in trying to remember, for example, a telephone number, we often rehearse it to ourselves by repeating it subvocally. That this procedure is important is clear if we attempt to engage our short term memory whilst repeating some other utterance (such as “blah blah blah”) either out loud or to ourselves. Our short term memory is thereby drastically impaired. (Baddeley calls this the ‘articulatory suppression’ of the ‘articulatory loop’.) The capacity that this procedure exemplifies Baddeley calls the ‘inner voice’. The ‘inner ear’, by contrast is not implicated in short term memory; it involves, for example, the ability to make rhyme judgements about visually presented words, an ability that is not impaired by ‘articulatory suppression’. Frith hypothesises that the presence of auditory hallucinations should impair short term memory tasks but not impair the making of rhyme judgements, and in making this hypothesis suggests that auditory hallucinations reflect a defect in the ‘inner voice’ and not the ‘inner ear’.

Given the empirical basis of Baddeley’s research, one might be tempted to infer that his theoretical categories (inner voice and inner ear) provide empirical evidence for the cognitivist’s conception of inner experience. That is, it could be argued that the possibility that the ‘articulatory loop’ may be suppressed is indicative of two aspects of inner experience (an inner voice and an inner ear), and trace a conceptual trajectory from the empirical evidence to the cognitivist’s metaphysical schema. But no such trajectory is indicated: Baddeley’s concepts of ‘inner voice’ and ‘inner ear’ remain picturesque – although theoretically unhelpful given the allure of the hopeless mentalism they encourage – remain picturesque labels for *just those very capacities he describes*. And these capacities are identified by the particular aptitudes shown in the experimental tasks, and not by reference to metaphysical concepts presupposed by cognitivist epistemology.

Baddeley’s concepts of the inner voice and inner ear differ from the cognitive theorist’s concept of a self-monitoring system or from the cognitivist’s concept of an inner sense in the following way: For the cognitivist what the ‘inner ear’ listens to *is* the ‘inner voice’: the inner ear (i.e. the self-monitoring system or the faculty of inner sense) signifies a passive and receptive capacity, the inner voice an active generative mechanism which generates that which the inner ear ‘introspects’. In Baddeley’s scheme, however, the ‘inner voice’ and ‘inner ear’ are just labels for two *quite separate* capacities (or perhaps for

the neurological mechanisms underlying the two capacities); Baddeley's inner ear does not listen to his inner voice.

None of this is to say that Frith's hypothesis - that schizophrenic hallucination involves Baddeley's inner voice - is in any way out of order. If it is true then such hallucination should impair short term memory but not the ability to make rhyme judgements. And if we are prepared to individuate psychological faculties or functions in the way Baddeley proposes then Frith's 'output' theory is a genuine psychological theory of the phenomenon of hearing voices. Given that faculties and functions so individuated are likely to owe their distinctness to discrete neurological systems, the output theory, if true, is a genuine advance in cognitive *neuropsychology*.

At this point in the elucidation of the theory, however, we are a long way from the initial characterisation or the development of this in the theory of 'self-monitoring'. The idea that inner speech is to be considered a self-initiated action which is not so recognised cannot itself be captured by the concepts of 'inner voice' or 'inner ear'. Only if Baddeley's cognitive-neurological theory is mischaracterised using the cognitivist's epistemological rhetoric of introspective access, inner sense, self-monitoring (in a psychological rather than neurological sense) etc. would it be thought to support the output theory in the general form in which it is developed by Frith. Even with the support of Baddeley's neuropsychology Frith's output theory of schizophrenic hallucination remains highly underdeveloped; without the 'support' of the cognitivist's dubious first-person epistemology the output theory consists of: an affirmation (contra the 'input' theory) of the true hallucinatory nature of hearing voices, along with a grammatical rather than empirical reminder of what it is to hear voices (to talk to oneself without realising it), and a suggestion that hearing voices may be due to an impairment in a certain functionally identified neurological mechanism (the 'inner voice'). As with their account of disorders of thought, language, and affect in schizophrenia, the cognitive theorist's theory of hearing voices only appears to be a *psychological* theory because of the misguided cognitivist epistemology that underpins its self-presentation. Translate the theory out of the cognitivist idiom and only a neurological mouse comes forth.

### **iii. Cognitive Theories of Thought Insertion**

Frith's self-monitoring theory of hallucination is intended to extend also to cases of thought insertion and other passivity experiences. Here is what Frith says about thought insertion (pp. 80-81):

Thought insertion, in particular, is an experience that is difficult to understand. Patients say that thoughts that are not their own are coming into their head. This experience implies that we have some way of recognising our own thoughts. It is as if each thought has a label on it saying "mine". If this labelling process goes wrong, then the thought would be perceived as alien.

This idea may sound fanciful when applied to thoughts. However there is ample evidence that such labelling does occur for various simple actions such as eye movements and limb movements [i.e. corollary discharge].

Our 'recognising [of] our own thoughts' is to be taken as an example of self-monitoring. This account of thought insertion is remarkably like that of G&S<sup>159</sup>. The fact that schizophrenic patients report that thoughts occur in their mind which are not their own is taken to indicate that we normally *do* recognise thoughts *as our own*, and that the schizophrenic's recognition faculty has become faulty.

Three arguments back up the theory as Frith presents it. Firstly, the fanciful idea that thoughts are *labelled* with information concerning their origin is argued to be less fanciful once we consider that a feedback loop exists in the visual system. The role of this comparison however is unclear. For the fancifulness of the idea of *thoughts* being labelled is not a general anxiety about *anything* being labelled, and even when talk of 'labelling' is translated into talk of corollary discharge, the difficulties in transposing the theoretical structure from a physiological to a psychological level of explanation are not diminished. (To modify an example from Chomsky: the plausibility of 'green ideas sleeping furiously' making sense is not increased by comparing it with 'black sheep sleeping peacefully'; what the comparison highlights is the semantic contrast, not the semantic similarity).

A second argument is given in a slightly later passage (p. 81), which could also plausibly be read as a clarification of the labelling idea.

I have suggested previously ... that a failure to monitor intentions to act would result in delusions of control and other passivity experiences. Thinking, like all our actions, is normally accompanied by a sense of effort and deliberate choice as we move from one thought to the next. If we found ourselves thinking without any awareness

---

<sup>159</sup> But n.b. Frith's account predates that of G&S.

of the sense of effort that reflects central monitoring, we might well experience these thoughts as alien and, thus, being inserted into our minds.

It is not that thoughts are actually labelled, then, but rather that we are either aware or unaware of the 'sense of effort' occurring when we think. This however substitutes phenomenological implausibility for metaphysical bizarreness, for there is in fact *very little* by way of 'a sense of effort and deliberate choice as we move from one thought to the next', especially when what is being considered are everyday rather than philosophical or deliberative thoughts. (Patients with thought insertion are no more prone to report philosophical rather than everyday thoughts inserted into their minds). And if we found ourselves thinking effortlessly, it would be most natural to express this experience just in these terms: we find *ourselves* able to think fluidly, with facility and without obstruction; and not that: *someone else* must be doing our thinking for us.<sup>160</sup>

A third characterisation of the defect in self-monitoring which could underlie the experience of thought insertion is also provided (p. 81):

intentions are monitored in order to distinguish between actions caused by our own goals and plans (willed actions) and actions that are in response to external events (stimulus-driven actions). Such monitoring is essential if we are to have some awareness of the causes of our actions. Given (as we have seen in Chapter 4) that different parts of the brain are concerned with willed action and with stimulus-driven action, this distinction could be made simply on the basis of which brain system was active.

As with the output theory of hallucination, it is possible to extract from the above a coherent empirical speculation concerning thought insertion, albeit one with a purely neurological content. If we follow this option, then *what it means* to talk about the monitoring of actions caused by our own goals or actions caused by external events is to talk about the neurophysiological feedback mechanisms that allow for a differential neurological response to movements which have their causal origins in different parts of the brain. But what *this* means is that self-monitoring theories of thought insertion just aren't psychological in character. Furthermore it is hard to see how the self-monitoring theory of action

---

<sup>160</sup> Frith now agrees that the 'sense of effort' idea is unclear. (In conversation.)

provides for a theory of thought insertion. Even if a physicalist theory of thought were accepted (in which thoughts or thinking were somehow 'identified' with brain states or processes), we do not (normally) first have intentions to think which are then followed by the thought in question any more than our thinking of a thought involves two components (a thought and the thinking of it) or an action on a content.

The difficulties involved in construing thought as an action have already been discussed in section 2 above. At this point however a further objection against both the philosophical account of thought insertion provided by G&S and Frith's psychological theory can be noted. On both stories the patient suffering thought insertion comes to their delusional belief that thoughts are being inserted into their minds because a mechanism which normally functions to allow the identification of one's thoughts as one's own has malfunctioned. The question of the relevance of this mechanism must however be doubted. What evolutionary advantage would such a mechanism provide for an animal? And what purpose would it serve? If animals did find themselves in the predicament of wondering whether a thought that they found in their mind was their own, the best way of getting them out of this dilemma would not be the provision of a mechanism for assessing whether or not the thought corresponded with some putative intention to think it (or what have you), but rather the kicking in of a mechanism that simply consistently provided a positive answer without checking. For it seems to be a simple *fact* that *all* the thoughts we do find in our heads *are* (in the relevant sense) our own<sup>161</sup>. In attempting to make sense of thought insertion by supposing that the subject with thought insertion is coherently interpreting an abnormal experience, the cognitive theorists not only fail to take stock of the delusional, psychotic nature of the symptom but also postulate 'fundamental' cognitive mechanisms which have no evolutionarily explicable purpose.

#### **4. The Alienated Roots of Cognitive Theories of Introspective Alienation**

---

<sup>161</sup> If telepathy had been a frequent occurrence in our evolutionary history the relevance of a self-monitoring system for our thoughts might be intelligible; it is however hard to imagine many cognitive theorists resting their theory on an ad hoc parapsychological speculation.

In section 2. ii. above I suggested that cognitive theories of thought insertion and auditory hallucination rely upon an alienated phenomenology for their depiction of our everyday psychological self-ascriptions. This mis-depiction of our normal capacity to avow our own thoughts and feelings as grounded in an *introspective function* makes way for the possibility of error in self-ascriptions. Thoughts for example are viewed as inner goings on - as mere *events* 'in consciousness' - to which we stand in some kind of *relation*. The contents of the mind being objectified in this way, the true self *retreats behind them*: our thinking itself becomes a two-fold matter of the active production of thoughts and the passive introspection of the same. If this is how our *ordinary* thinking is viewed, the project of understanding thought insertion is greatly simplified: the delusional patient's reports that thoughts are being 'put into their minds' which are not their own is straightforwardly intelligible given the cognitive malfunction mooted: the self-monitoring of the active production of thoughts breaks down whilst the passive introspection of thoughts remains intact.

On this cognitive formulation the experiences of 'introspective alienation' are (putatively) made intelligible as malfunctions within a normal faculty of self-consciousness. By contrast to these cognitive formulations, the writings of the phenomenological psychologist Louis Sass can be viewed as providing a radically alternative model. On his scheme, the experiences of introspective alienation (thought insertion etc.) is not to be understood as a disorder within a normal faculty of self-consciousness, but rather as the product of the intrusion of self-consciousness into a mind which would be healthier without it. On my reading, the cognitive theorist is in error in taking an alienated and objectified view of the character of mind; on Sass' reading, it is the subject with schizophrenia that is to be viewed as taking an introspective ('hyper-reflexive') alienating and objectifying attitude toward the contents of their own mind. Where on my scheme the *cognitive theorist* makes an error in their theorising, on Sass' scheme the *subject with schizophrenia* embodies this error. They do not themselves make any kind of *mistake* in their thinking; rather, the *form* of their thinking itself *incarnates* the alienated perspective of the cognitivist.<sup>162</sup>

Sass' scheme is more than ingenious and the analogy with cognitivism (or 'modernism' as he calls it) is illuminating. The analogy 'makes sense of' schizophrenic experience in as much as it provides a

---

<sup>162</sup> The embodiment of this objectifying mentalistic conception of consciousness is only half of Sass's story; the other half involves (as a kind of flip-side) the subjectivising within schizophrenic consciousness of the objective external world.

unifying framework within which a whole variety of psychotic experience can be placed. The manner in which it makes sense of schizophrenia is however very different from that of the cognitive theorist. The cognitive theorist aims to explain by providing causal explanation involving defects in underlying mechanisms; in reality the psychological (rather than neuropsychological) character of such explanations works by casting what the subject with schizophrenia says and does as a kind of intelligible *reaction* on their part (the kind of thing an otherwise non-psychotic subject might think) to the cognitive malfunctions within them. Sass by contrast explicates schizophrenia not by providing causal explanation but by using an analogy to *situate* the various psychotic manifestations within an overall scheme.

What is particularly striking but also somewhat theoretically disturbing about Sass' scheme is that it self-consciously deploys, as an analogy to provide understanding, an ultimately unintelligible metaphysics and epistemology. The subject with schizophrenia is viewed as embodying not merely an *unusual* form of consciousness, but rather as instantiating a *logically impossible* form of mindedness. On the one hand this is a strength of the theory: the logical instability of the mechanistic mentalist's conception of mind is to be understood as an analogy for the inherent psychotic instability of the schizophrenic mind. On the other hand however it leaves unanswered the central psychological question of what it means to be 'schizophrenic', for schizophrenia is after all not a logically impossible phenomenon but an all-too-common empirical condition.

I shall return to the question of how to satisfy the demand for psychological explanation of thought insertion in particular and delusional symptoms more generally in the last chapter. For now I wish merely to recap what has gone wrong in the cognitive theorisations of thought insertion and auditory hallucination, and to take some of the pressure off the critic of such cognitive theories, pressure prompted by very natural intuitions that *something* akin to the explanations on offer must be correct and that some such psychological question about why a patient suffering from schizophrenia believes that thoughts are being put into their head demands an answer. In fact and of course I do not wish to argue that no kind of psychological understanding of thought insertion can be provided. The issue rather turns on the *kind* of explanation required. In line with the critique of cognitivist epistemology developed in the first two parts of this dissertation, the line pursued here will be that it is only a presupposed alienated conception of mind that generates an apparent need for the kinds of epistemological and psychological explanations

that the cognitive theorist aims to provide. Doing away with mentalistically motivated epistemologies does not however leave us completely in the dark about the psychopathological phenomena: in coming to understand the redundancy of the cognitivist questions our understanding of the nature of the mind and of its disorders is simultaneously enhanced.

Three factors conspire in generating the perceived need and appearance of propriety of cognitive theories of introspective alienation. First there is the psychopathological phenomenon itself: the subject with schizophrenia reports having thoughts which are not their own inserted into their heads, a description which is most naturally couched within a mentalistic theory of consciousness: thoughts as inner objects which can be brought into view within a passive introspective faculty (but of which we might remain unaware) and which are produced by active generative acts of thinking. Secondly there is the mentalist's alienated conception of the mind itself. And finally there is the provision of alienated phenomenologies of thought and inner speech, descriptions of supposedly normal and also abnormal thought which maintain a highly disengaged stance toward the phenomena and which serve to render plausible the cognitivist theorisation of the phenomena.

All three influences can be seen in the following example taken from Frith's theory of introspective alienation in terms of a failure of 'metarepresentation'. Frith has, as noted above, now abandoned the idea that this theory could plausibly unify his failure in self-monitoring and failure in willed action theories of schizophrenia<sup>163</sup>, but the example is nevertheless instructive (pp.125-6).

... metarepresentation, in the sense in which I use the term, is concerned with knowledge like "Mary believes 'John is sad'" or "Mary believes 'bananas are yellow'". Thus metarepresentation is concerned with knowledge about representations. This knowledge will have two components, the form of the representation and its content. For example "I know 'X'", "Mary believes 'X'", "I intend 'X'", all have different forms, but the same content (X). I propose that in some schizophrenic patients metarepresentation fails in such a way that the patient remains aware only of the content of these propositions. Thus I (Chris) might infer about my friend Eve the proposition, "Eve believes 'Chris drinks too much'". If my mechanism for metarepresentation failed, then, when I thought about Eve, the free floating notion "Chris drinks too much" might enter my awareness. If I described this experience it would be called a third person hallucination.

---

<sup>163</sup> In discussion. See also p. 133 where the theory is modestly described as 'doubtless over-inclusive'.



First a psychologistic conception of belief is invoked. Chris' belief that his friend Eve thinks that he drinks too much, which on a natural enough analysis of belief might thought to entail certain *hypothetical* attitudes toward propositions (if asked whether the proposition 'Eve thinks Chris drinks too much' is true, Chris should respond in the affirmative if he truly believes what the proposition asserts), is taken to involve an actual attitude toward an actual proposition. This proposition therefore requires psychological realisation in Chris' mind. For Chris to have the true thought that Eve believes he drinks too much both Chris and Eve must mull over certain propositions in their heads. (For this reason Chris' belief that Eve thinks he drinks too much becomes an *inference* – to something (inner speech) hidden and inner, and what is inferred is a *proposition* – 'Thus I (Chris) might infer about my friend Eve the proposition, "Eve believes 'Chris drinks too much'".'))

The idea that belief involves propositional content in this psychologically real way underpins the otherwise unclear idea that metarepresentation (the capacity to have thoughts about those mental contents of others which are in turn concerned with the mental contents of others) could plausibly be thought of as dependent on a single *mechanism*. (If having a belief about a belief about a belief involves putting various propositions through their paces it is all too natural to suppose that a certain mechanism to effect these transformations will be required. If having a belief involves no such psychological processing of inner representations then no such mechanism will be needed.)

The sneakiest part of the theory which quickly covers over its incoherences is found in the suggestion that when Chris thinks about *Eve*, the free floating notion 'Chris drinks too much' might enter his awareness. Why this should be so is not clear: Chris presumably has thousands if not millions of beliefs about Eve. What the example really requires is the suggestion that when Chris thinks about *Eve's belief that he drinks too much*, the free-floating proposition 'Chris drinks too much' would enter his consciousness. But the reason why this most natural way of spelling out the theory has to be avoided is that it takes away with one hand what it gives with the other. On the one hand it wants to suggest that Chris has a certain thought, but on the other it wants to deny this: what Chris thinks is not that Eve thinks that he drinks too much, but that Chris drinks too much.

The difficulty in sustaining any such account comes from the difficulty of supplying ascription conditions for the thought which on the one hand needs to be ascribed to Chris for it to be said that his

metarepresentational system fails but which on the hand needs to be denied if what Chris actually thinks is that Chris drinks too much. The avowalist treatment of thought presented in chapter 3 provides us with a clear criterion for the content of thought: what someone thinks is what they are sincerely prepared to say that they think. The cognitivist however can provide no such criterion, and is forced to leave us with the pictorial suggestion that the content of someone's thought is given by whatever it is that the subject 'represents' to themselves. In this way the constitutive subjectivity of the mental is denied and replaced with an objectivising and alienated conception of mind; but without any means of interpreting this picture we are at a loss to know what to do with it.

Once again the temptation is surely to suppose that, whatever the failure in the details, the fact of the existence of thought insertion entails that some such explanation must be in order, that some account of the structure of the mind which makes way for the possibility of error in identifying the origin of our thoughts (as in G&S's example) or error in accessing the content of our thoughts (as in Frith's example) is required. Whether or not the rejection of this demand can be sustained will have to wait until chapter 7, where it will be argued that the reports by subjects with schizophrenia that thoughts are in their heads which they do not recognise as their own can be taken seriously without being taken literally, without implying the truth of the cognitivist's theory of thought.

To finish consider the kind of explanation that G&S wish to provide for thought insertion. Whilst it might be thought that something like this kind of explanation *must* be right, I shall argue that the actual details reveal that *no* explanation of this sort *could* be right. What G&S ask is whether (p. 97<sup>164</sup>) we can 'provide a coherent and plausible interpretation of the subject's assertion that an episode occurring in his mind is attributable to someone else rather than to himself.' And they argue that (p. 92) 'the key to making sense of such reports, the key to making them coherent, lies in recognising a distinction between attributions of subjectivity and attributions of agency.' This enables them to suggest that the beliefs of the schizophrenic concerning the operation of their own mind (p. 93) 'represent, for the most part, honest errors'. As they say in another context<sup>165</sup>: 'Our case will require arguing that delusions of thought insertion are, in some respects, less bizarre and more coherent than they otherwise appear.'

---

<sup>164</sup> Page references in what follow are to *Mind and Mine*.

<sup>165</sup> *When Self-Consciousness Breaks*, p. 10.

Not only do the distinctions drawn by G&S make for a suspect conception of subjectivity, but what the distinctions enable us to do is to 'make sense' of thought insertion only by depriving it of its status as a paradigmatically schizophrenic symptom. The patient's ideas of thought insertion are seen as plausible and coherent, as honest errors. But this is precisely what reports of thought insertion are not: thought insertion is classified as a delusion and not as a mere experiential disorder for precisely the quality of psychotic strangeness that belongs to it. G&S gloss 'the key to making sense of such reports' as 'the key to making them coherent', but what is really required is a way of coming to understand and in this sense 'make sense of' thought insertion which *isn't* forced to make it coherent. To deny that the form of explanation provided by G&S is apt is not to forego the important attempt to achieve empathy with and understanding of the schizophrenic condition; it is rather to redefine the very phenomena understanding of which is required.

In retrospect it is clear why G&S are drawn to the idea that the way to make sense of psychotic phenomena is to make the phenomena themselves make sense. For the most striking thing about their general scheme is the way in which the true self retreats behind its own mind, becoming both an inner producer of the thoughts that are to be found there and an inner observer of the same. As remarked in Part 2, the self that disengages from the world, retreating inside the body and behind the mind, is a fully *sane* self, equipped with all of those faculties (the capacity to engage in inner actions, the capacity to perceive and make sense of inner representations etc.) that the actual self *qua* person (embodied subject) possesses, faculties which the cognitive models are supposed to be explaining. The cognitivist model of the rational agent tacitly presupposes an internal disengaged self to supervise the rational functions of the engaged subject. The cognitivist model of the psychotic subject similarly presupposes that behind the appearance of a disordered mind lies a rational subject doing the best of a bad job. It is for this reason that the two models sit together so neatly. But as long as the cognitive models must presuppose rather than theorise our rational contact with the world, or our capacity to self-ascribe thoughts, or our capacity to act intentionally, then they will be unable to theorise psychotic illnesses involving fundamental disturbances in the self, in contact with the world, and in rational action.

The last two chapters have considered cognitive theories of schizophrenic disorders of action, language, disordered thought, the auditory hallucination of voices, and thought insertion. Essential to any true understanding of schizophrenia however, and essential even for understanding what it is that makes the above-mentioned symptoms *schizophrenic* symptoms, is an appreciation of what it means to be *deluded*. The next chapter (6) contains a critique of cognitive theories of delusion; the last chapter (7) develops an alternative descriptive phenomenology of delusion which descriptively extends to the cognitive and experiential symptoms considered in chapters 4 & 5. In line with the approach developed in Part 1, the intent will be to supply an initial description of the phenomena which casts them in such a light that the kind of explanatory projects encouraged and envisaged by cognitivism appear as pressing or desirable.

## **Part 4**

# **The Psychotic Core**

## Ch. 6. Cognitive Theories of Schizophrenic World-Estrangement

### 1. Introduction: The Phenomenology and Epistemology of Delusion

The symptoms so far examined – the experience of thought insertion, thought and language disorder, poverty of action and auditory hallucination – are all important in the diagnosis of schizophrenia, but none are so important as delusion itself. Indeed, hallucinations are most indicative of psychosis when they are *delusionally believed to be real*, and the same goes for passivity experiences, which in schizophrenia are likely to be *delusionally* understood as indicating the influence of an outside agency. As Kemp et al say, delusions ‘are the *sine qua non* of psychosis’<sup>166</sup>, and it is through a consideration of delusion that we are able to get to the core of schizophrenia and come to an understanding of what it means to suffer from this illness.

Anyone trying to gain a psychological understanding of delusion today would naturally turn to the cognitive sciences, concerned as they are with that – thought, belief, understanding, rationality – which is apparently disturbed in delusion. And, as shall be seen in what follows, various cognitive psychological theories of delusion have indeed been proposed. What this chapter will argue, however, is that such theories are actually not at all helpful in understanding the most prototypical and most ‘delusional’ of schizophrenic delusions, and that this is due to a double mutually supporting failure. On the one (phenomenological) hand insufficient attention is paid by cognitive theories to the actual character of genuine delusions; on the other (epistemological) hand cognitive psychological theories tend to inherit from philosophical cognitivism certain dubious assumptions about our contact with reality and the foundations of our rational life. What the cognitive theories of delusion end up providing are coherent enough theories of how certain *unusual* or *false* beliefs may come about, but these beliefs are *not delusions*. They do not demonstrate the fundamental irrationality and lack of contact with reality which gives delusion its central place in the diagnosis of psychosis. This, I shall argue, is not surprising since

---

<sup>166</sup> *Reasoning and Delusions*, p. 398.

cognitivism does not provide the epistemological resources with which to adequately theorise our contact with reality in the first place.

In pursuing the question of delusion it might be found tempting to separate out the issues into two categories: one concerning the *meaning* of 'delusion' (known in the literature as *das Wahnproblem*), the other concerning the typical *causes* of delusions in schizophrenia. The causal rhetoric of cognitive theories of delusion in particular might give the impression that its empirical theories are effectively insulated from the kind of analytic investigation of the meaning of delusion elaborated in this (and in more detail in the following) chapter. This however would be misleading. First, it is possible that the causal rhetoric which dresses up the cognitive theories is misleading as a guide to the kind of explanatory work that such theories do. For example, a cognitive theory which hypothesises that delusions are *caused* by abnormal experiences might really be interested in whether delusions can be *made intelligible* to us given such and such experiences, and an examination of how the explanations are *made* (rather than of how they are *portrayed*) might reveal a concern with situating delusions in the *logical space of reasons* than in the *realm of 'natural law'*. (The delusion being made explicable in terms of the canons of rationality, not in terms of causal antecedents.<sup>167</sup>) Secondly, it may be that the analysis of 'delusion' reveals that certain types of explanation can be ruled out *a priori*. Perhaps, for example, delusions are to be understood negatively, as that which cannot be explained in rationalising terms, or even functionally, as types of thoughts which have such and such (but not other) typical causes. In what follows I shall suggest that reflection on what it means to be deluded reveals that common cognitive theories (however interpreted) simply *cannot* constitute adequate explanations of delusion.

In coming to understand what is meant by 'delusion' it will be useful to introduce certain concepts from psychopathology, in particular from Jaspers' *General Psychopathology*. One key distinction is between the *internal* and *external* characteristics of delusion, a classification which distinguishes the *essential* (internal) core features of delusion and delusional thought from (external) characteristics or

---

<sup>167</sup> This is not to suggest that causal and rationalising concerns exhaust the scope of the concept of making things intelligible. As I shall suggest below, delusions (by definition) cannot be understood as rational responses, and causal concerns are not necessarily relevant; instead, *phenomenological description* of the general characteristics of delusion provides the understanding required to make psychological sense out of them.

concerns which delusions typically manifest but which are not specific to or essential for delusion itself. (Internal features of delusions relate mainly to what is known as their *form*; external features by contrast make reference to their *content*, or to merely extrinsic features of their form.) Another psychopathological distinction is that between *primary* and *secondary* delusions, also known as the distinction between *true* delusions and *delusion-like ideas*. As the simple version of the story goes, primary delusions are ununderstandable and cannot be reached through empathy, whilst secondary delusions are rationally intelligible given the mood state or the past or present experiences of the subject<sup>168</sup>.

This latter distinction needs some more careful spelling out, and both concepts will be examined further in section 2 below, but their relevance to the discussion of cognitive theories of delusions can already be anticipated. Cognitive theories, it will be argued, fail to get to the core of the issues, fail to plumb the depths of what it means to be deluded, and part of the reason for this is that they implicitly take up a stance on the psychopathological issues just raised. In their characterisation of what it is that they investigate, the theories rely on descriptions of delusion that invoke only external characteristics, and it is therefore doubtful whether they address themselves to the real disturbance. Furthermore, they commonly suppose that it is a merely empirical issue whether the most characteristic of schizophrenic delusions are primary or secondary, and because they aim to make such delusions intelligible by giving (a particular form of) psychological explanations for them, typically presuppose that genuine delusions are secondary.

Such presuppositions concerning psychopathology dovetail nicely, it will be argued, with certain *epistemological* assumptions encouraged by cognitivism, and it is because these presuppositions and assumptions fit together nicely and allow a story to be told which has an internal coherence to it that the deficiencies of the resulting cognitive theories frequently go unnoticed. The two principle such assumptions embodied by cognitivism are: i) that the foundation of our *contact with reality*, our epistemic engagement with the world, is structured first and foremost by *knowledge and belief* (or *mental representation*), ii) that the bedrock of our *rationality* is to be discerned in our correct *reasoning*, in our

---

<sup>168</sup> For elaboration of the distinction see Andrew Sims' *Symptoms in the Mind* pp. 104 ff. Jaspers discusses delusion in *General*

*Psychopathology* (henceforth *GP*) part 1, chapter 1, section 1, subsection §4.



accurately going from premise to conclusion in deductive and other *inference*. Such epistemological assumptions, it will be suggested, involve an alienated and topsy-turvy conception of the place of the subject in the world, and a disengaged conception of rationality. The active participation and know-how of the agent is (impossibly) made consequent upon their knowledge, and the coherence of action and communication is made (again impossibly) consequent upon what is in fact an abstract mode of thought (logical reasoning).

The dovetailing of the philosophical and psychopathological assumptions, then, works in the following way. Cognitive theories characterise delusions in terms of their external characteristics and also typically view them as secondary, as rationally intelligible responses. This fails to access the depths of the delusional condition. And cognitivist epistemology fails to correctly characterise the foundations of our rational engagement with the world. But the shallow conception of *that lack of contact with reality constitutive of the psychotic condition* provided by the cognitive characterisation of delusion looks as if it could readily be accounted for in terms of the shallow conception of our *contact with reality provided by the cognitivist epistemology*. If delusions indicate failures in the content of knowledge, or indicate a failure to preserve the truth-values of such contents in propositional reasoning, and if our worldly orientation is underwritten by true beliefs and correct reasoning, then it would of course be natural to understand a lack of contact with reality in terms of a failure of knowledge or of inference. But, the argument will be, both the epistemological and the psychopathological positions are false. Delusions are more deeply strange than the cognitive psychologist admits, and the grounds of our contact with reality are not to be found in mere representational knowledge, but rather in the depths of the agential Background.

In previous chapters the argument has all been from the side of philosophy, but in what follows I shall also argue from a psychopathological point of view that the characterisation of delusion on offer fails to cohere with what we know of the diagnostic practice of mental health practitioners and with our common intuitions about what does and what does not count as 'mad'. In the next chapter (7), I shall suggest how a re-worked epistemology will enable us to characterise delusion in a positive way, and thereby come to an understanding of what it means to be psychotic.

## 2. Preliminary Remarks on the Character of Delusion

### i. Internal and External Characteristics

Many cognitive theories of delusion start with a definition of the term, usually one based on DSMIII, IIR or IV<sup>169</sup>. But whilst it might be natural to suppose that the diagnostic manuals should be considered *prescriptive* for diagnosis and therefore (logically) impervious to error, they are in fact more helpfully thought of as *descriptive* of the grounds of diagnostic practice.<sup>170</sup> This then leaves open the logical possibility of error, and also makes room for the possibility that psychological theorists, relying more on the common diagnostic manual understanding of delusion than on a clinical pre-understanding of the concept, may choose explanatory constructs that fail to do justice to the true character of the explanandum. In what follows I shall argue that the textbooks fail in their definitions of delusion, and this because they identify delusions solely in terms of their extrinsic or 'external' features. Later I shall suggest that cognitive theories of delusion tend to address themselves to beliefs which have the external

---

<sup>169</sup> DSM is the *Diagnostic and Statistical Manual of Mental Disorders* of the American Psychiatric Association. For examples of cognitive theories employing DSM or closely related definitions, see Philippa Garety, *Reasoning and Delusions*, p. 14; Kemp et al. *Op Cit.*, p.398; Brendan Maher, *Anomalous Experience and Delusional Thinking: The Logic of Explanations*, p. 15;

<sup>170</sup> This perhaps especially so when considering mental as opposed to physical illness. The capacity to accurately diagnose mental illness is a *practical skill*, picked up after considerable involvement and experience both with patients and with trained psychiatrists; it is not something that can be learned simply by reading a diagnostic manual; the trainee requires to be given examples and not merely rules. It is not surprising then that, as we shall see, the 'definitions' of delusion given in the textbooks rarely square with actual medical practice. (For more on this see Jeff Coulter's *Approaches to Insanity* chapters 1 & 4, and also K W M Fulford's *Moral Theory and Medical Practice*, chapter 10, which exposes various discrepancies between the diagnostic activities of psychiatrists (vis-à-vis 'psychosis' and 'delusion') and the descriptions and recommendations of the diagnostic manuals) (This is not to say that we should be distrustful of the manuals' and textbooks' definitions of *mental illnesses* – of the criteria that they provide for 'schizophrenia' or 'bipolar disorder' for example – although the arbitrariness of some of the (sometimes highly disjunctive) criteria for some such may disincline us from making too much of such schemes. (At least the standardisation of diagnostic criteria for mental illnesses has largely resolved the widespread inter-practitioner disagreements of the pre-1970s over, for example, the diagnosis of schizophrenia.) Such definitions of mental illnesses however presuppose a mastery of the correct use of 'delusion', 'hallucination' etc., which is to say, a mastery over the diagnosis of delusion and hallucination. It is in the definition of such key psychopathological concepts that, arguably, the manuals fall far short.)

but not the intrinsic or 'internal' characteristics of delusion, and thereby fail to adequately address the psychopathology in question.

When Karl Jaspers first attempted a description of the concept of delusion<sup>171</sup>, he provided a list of what he termed 'mere external [or extrinsic] characteristics':

The term delusion is *vaguely* applied to all false judgements that share the following external characteristics to a marked, though undefined, degree: (1) they are held with an *extraordinary conviction*, with an incomparable, *subjective certainty*; (2) there is an *imperviousness* to other experiences and to compelling counter-argument; (3) their content is *impossible*.

As well as noting that this is only a *vague* definition based on *merely external* characteristics, Jaspers goes on immediately to call into question the validity of this definition *qua* solution to *das Wahnproblem* (GP p. 93):

To say simply that a delusion is a mistaken idea which is firmly held by the patient and which cannot be corrected gives only a superficial and incorrect answer to the problem.

Notwithstanding his reservations, it is clear that it is from such avowedly deficient external characteristics that many current definitions take their lead. DSMIIIR, for example, defines a delusion as

a false personal belief based on incorrect inference about external reality and firmly sustained in spite of what almost everyone else believes and in spite of what constitutes incontrovertible and obvious proof or evidence to the contrary. The belief is not ordinarily accepted by other members of the person's culture or subculture (i.e. it is not an article of religious faith). When a false belief involves an extreme value judgement, it is regarded as a delusion only when the judgement is so extreme as to defy credibility.

The authors of the latest DSM (IV – 1994) seem to have pretty much given up on the project of defining delusions<sup>172</sup>, but many other examples could be given – I shall include just one more, from Leff and Isaacs' handbook of *Psychiatric Examination in Clinical Practice* (p. 50):

---

<sup>171</sup> GP, pp. 95 ff.

A delusion is a false belief, firmly held by the patient, which is not consistent with the information available to him and with the beliefs of his cultural group, and which cannot be dispelled by argument or proof to the contrary.

Correcting these standard definitions with an alternative characterisation will be the project of chapter 7; for now it is sufficient to note the failures of the definitions. Consider first the view that delusions are necessarily *false* or *impossible* beliefs: we are immediately confronted by the possibility that a delusion manifest in 'an extreme value judgement', being evaluative and not descriptive, might not be helpfully be thought of as false. Or if 'falsity' is to capture what it is about an evaluative delusion that reveals its wrong-headedness, we are left with needing to know what 'delusion' means *before* we can understand this particular psychiatric use of 'falsity'. Furthermore, delusions need not concern (logically?, empirically?) impossible situations: it is not impossible but only improbable that one's flat has been bugged by the CIA. Even more tellingly, some beliefs may be judged as delusional by a psychiatrist even without evidence of the falsity or truth of the belief, or *even when the belief is known to be true*. A Mr A.B., for example, was diagnosed as suffering from the hypochondriacal delusion that he had brain cancer. Now Mr A.B. did not in fact have cancer, but someone in his condition (depressed, suffering from headaches) *could* have had such a cancer. The relevant concern is that the diagnosis of delusion was not made pending the results of a brain scan, and *would not have been withdrawn even if the brain scan had revealed a cancer*.

This latter point is made more readily intelligible if we consider the diagnosis of the Othello syndrome, a condition of pathological jealousy in which the patient (usually a man) suffers the delusion that their partner is cheating on them. Even if not true at the time of diagnosis (although it may be), the syndrome can be so severe (in its social impact) as to become self-fulfilling. Consider too the patient diagnosed with the delusion that they were mentally ill. If the standard definition were true, we would be

---

<sup>172</sup> All that the authors say is that (pp. 274–5) the positive symptoms of schizophrenia include 'distortions or exaggerations of inferential thinking (delusions)', and that 'Delusions ... are erroneous beliefs that usually involve a misinterpretation of perceptions and experiences ... The distinction between a delusion and a strongly held idea ... depends on the degree of conviction with which the belief is held despite clear contradictory evidence.'

left with a paradox: the patient is correct if the diagnosis were correct, but this would imply that the belief in question were not false but true, in which case, on the definition provided by the manuals, the patient would not be suffering a delusion after all – in which case, their belief would be false. What such examples show quite clearly is that talk of falsity (or impossibility) is not apt for characterising the wrong-headed strangeness of delusions.<sup>173</sup>

Consider now the view expressed by DSM-III-R and IV that delusions are based on ‘incorrect [or distorted or exaggerated] inference about external reality’. It is not clear what is meant by ‘external reality’ in this context – most delusions for example (and as the definition notes) concern the patient themselves in some way, and often make reference to what is going on in their own minds (thought withdrawal etc.). It is also unclear that distorted and exaggerated and incorrect inferences are the preserve of the delusional population. For that matter, the ideas of the lone scientific genius, some of which even whilst incredibly innovative are likely to be false, and ahead of their time and so not shared by others, are hardly to be counted delusions simply on this basis. Furthermore the possibility of finding incontrovertible and obvious evidence or proof against the delusional belief is not always present (consider thought withdrawal again), but is unlikely to affect diagnostic practice. Most telling of all, many of the most prototypical schizophrenic delusions are autochthonous – arising suddenly and out of the blue – and perhaps are not therefore best thought of as based on any kind of inference. Indeed, in Jaspers’ presentation of the character of primary or ‘true’ delusions, when he continues from his initial presentation of ‘external characteristics’ to give some kind of account of the psychological core of delusion, he straightaway distinguishes delusions proper – the true ‘delusional experience itself’ – from elaborations of this delusion or from judgements based upon it (*GP* p. 96). It is questionable, then, whether delusions are based on *inferences* of any sort.<sup>174</sup> (I shall return to this issue in section 3 below.)

---

<sup>173</sup> The examples in this and the previous paragraph are taken from K WM Fulford’s *Moral Theory and Medical Practice*, pp. 201–205; cf also *GP*, p. 106.

<sup>174</sup> Except to the degree that any false beliefs (and delusions are usually false) can be said to be based on inference (given that they can hardly indicate a direct experience of the way things actually are). In this context, it would hardly be contrived to suspect the DSM definitions as implying a tacit subscription to the mentalist’s questionable epistemology: we form beliefs about the way things are in the *external* world by making *extrapolations* or *inferences* on the basis of *inner* experiences. (By way of resisting this picture, I should like to stress that I do not generally come to a perceptual belief *on the basis of* what I see or hear, if by this is

The above argument is based on the view that, as the diagnosis of delusion is to some extent based on a capacity to recognise a particular quality of oddity in certain beliefs that we can all exercise, and within psychiatry is grounded in practical training and engagement with patients and older and wiser practitioners, the definition of 'delusion' in the textbooks must be responsible to that which psychiatrists diagnose. And not, that is, vice versa.<sup>175</sup>

Consider finally, and as a way into understanding the meaning of the internal/external distinction, the DSM suggestion that delusions are 'not ordinarily accepted by other members of the person's culture or subculture (i.e. it is not an article of religious faith)'. What is implausible about this suggestion is not simply the fact that, for example, articles of religious faith may be non-delusionally held which are not representative of one's culture or even subculture. Rather, it seems more plausible to suggest that there is a reason why delusions are not accepted by members of one's subculture (family, friends etc.), which is simply *because they are delusions*, and the other members of the subculture are *not insane*.

To elaborate: What is surely wanted from a definition of 'delusion' is an explanation of what it is in which the particular and peculiar oddity of delusional beliefs consists. But what the manuals provide is something quite different. On the one hand they tell us that delusions are 'false' or 'incurable' or as 'extreme', but we have found that such descriptions are either just wrong in so far as delusions are not necessarily false, or misleading and uninformative in so far as a particular psychiatric meaning of 'falsity' must be imported which has yet to be explained. In other words, the manuals seem to trade on a pre-understanding of 'delusion' in terms of which its descriptive terms ('false' etc.) are to be understood; a situation which gets us nowhere with its circularity. On the other hand the oddity of delusions is interpreted *statistically* rather than *normatively* – we are told that delusional beliefs are odd (in the sense of culturally uncommon rather than 'crazy'); but the strong suspicion surely remains that this is just because *most people are not deluded*. The fact that terms such as 'unlikely' or 'odd' have a dual meaning – one referring to the *plausibility* of beliefs, the other to their *prevalence* – hardly helps the situation.

---

meant that I come to my belief that my cat is on my desk by inferring its presence from my seeing it there. Rather, I simply believe *what I see*; in seeing that the cat is on the desk, I believe – or more likely, I know – that it is there).

<sup>175</sup> If this is right then anti-psychiatric suggestions that talk of delusion is invalid, when they are based on a critique of the common definition (such as Mary Boyle's in *Schizophrenia: A scientific delusion?*, pp. 212–217), can be argued to be beside the point.

That delusions are odd (as in statistically unusual) in the sub-culture is an *external* characteristic of delusion; but their intrinsic peculiarity (when assessed against the canons of rationality) must be assessed using the *internal* criteria of delusion. (It is by conflating the statistical and the normative that psychiatric abuses of power can be spuriously legitimised, as were once found in the old Soviet Union, dissidents being sectioned and said to be suffering from 'reformist delusions'<sup>176</sup>.)

It might be thought that the problems encountered so far in the analysis result from a simplistic conception of meaning. What is admittedly being sought is the 'psychological *core*' of delusion, but it could be objected that in searching for the essence of delusion we ignore the possibility that 'delusion' is a 'family-resemblance concept'. This would be to suggest that all delusions may not share one definable essence or core, but may rather have overlapping properties, resembling one another like the members of a family.<sup>177</sup> The 'failure' of the standard criteria may not, then, really indicate a failure on their part, but rather a failure within essentialist theories of meaning. But even if it were true that delusion is a family-resemblance concept, the argument can still be made that the standard theories simply fail to characterise or explain the oddity of delusional beliefs and of the reasoning that they embody. To be told that delusions need not be false *and* unusual but could be false *or* unusual does not help at all in explaining the manner in which delusions can be thought of as 'false' or 'unusual'. The strong intuition is that the true nature of delusions has not been touched by the supposed criteria, and that even if *most* delusions are false, and *this perhaps no accident*, the appeal to falsity has not touched the incomparably strange nature of the delusion in which its delusionality consists.

Given that the distinctive peculiarity of delusional beliefs is not captured by the standard criteria, it may not be unreasonable to suppose that such criteria are really only empirical markers of delusion after all, and furthermore that 'delusion' may not after all be a family resemblance concept. The attempt should at least be made to see if a non-disjunctive definition of the term could be given, and the standard manuals have hardly made a start on this project, given that they only take up from Jaspers that part of

---

<sup>176</sup> Cf. Medvydev & Medvydev, *A Question of Madness*, cited in Maher *op cit.* p. 17.

<sup>177</sup> The 'family resemblance' concept is Wittgenstein's. The literature on the related topic of (psychotic loss of) insight sometimes makes use of a similar strategy: it suggests that failures in the attempt to define 'insight' categorically can be vitiated by viewing insight instead as a *dimensional construct*.

his account which he explicitly avows to be inadequate, and totally fail to consider his more detailed pronouncements. (Only if this attempt proved ultimately unsuccessful would we be justified in opting for a disjunctive definition.)

To return now to the question of the internal/external distinction: Striking about the standard definition of 'delusion' is that it relies to some degree on identifying delusions with reference to their *content* and not their *form*. True, the content just in itself is not considered relevant, but it is with reference to the content that the diagnosis is largely supposed to be made. It is the content that must be considered, for example, when an assessment of the (supposedly prerequisite) *falsity* of the belief is made, or when a statistical evaluation of the *prevalence* of the belief in the subculture is carried out. And it is such markers of delusion that have been argued to be the most unreliable. The form is considered at other times (a delusion is considered necessarily *firmly sustained in the face of counter-argument or evidence to the contrary*), but with little detail<sup>178</sup>. This however must be a mistake. For a psychotic subject has lost touch with reality, lost contact with the world, and it is in their delusions that this lack of contact is most readily apparent. It is in what we might call the *structural rôle played by a delusion within the cognitive economy of the subject*, or the manner in which delusion represents a *failure of the grounding of the subject in reality*, that its essential characteristics must be sought. Such aspects of delusion clearly relate to its form, and they are to be considered the internal characteristics of delusion. A theory which fails to address itself to beliefs with such an epistemic status can hardly be a theory of delusion, and this will be a risk run by all psychological theories which identify delusions solely by means of external, derivative characteristics.

## ii. Primary and Secondary Delusions, Or 'True' Delusions and 'Delusion-Like Ideas'

---

<sup>178</sup> P. Mullen in *The Phenomenology of Disordered Function*, provides the following remark on the form of delusion (p. 29): 'The precision of the distinction between form and content possible in disorders of perception [e.g. hallucination] is less easy to maintain in disorders of belief. A delusion is nevertheless judged to be present more from the manner in which it is adhered to and the reason for its emergence than any aspect of content.' (The section on delusion ends thus (p. 40): 'Delusion represents a profound and complex disorganisation of mental life stretching way beyond mere false ideas and mistaken beliefs.')



It is pretty much a commonplace today to find, in psychiatric and psychological texts, the lament that Jaspers' definition of true delusions as 'ununderstandable' and 'psychologically irreducible' takes up a rather depressing, unhelpful and unhelpful view on the nature of delusion, and decides the issues in an *a priori* rather than (a more proper) investigative manner. For example:

The psychiatrist and philosopher Karl Jaspers said that the ideas of the person with schizophrenia are intrinsically non-understandable: they are simply the by-products of a biological disorder, lacking all rational structure.<sup>179</sup>

To some extent, this prejudice is consistent with Jaspers' influential view that psychotic symptoms are 'ununderstandable' and do not reflect patients' personalities and experiences.<sup>180</sup>

Jaspers argued that the symptoms of all [non-schizophrenic] mental illness were comprehensible as exaggerations of normal states of anxiety, euphoria, depression, fear, grandiosity, and the like. ... Contrast these with the sheer strangeness of ... schizophrenic experiences...., with their uncanny mutation of all normal relationships between self and world. ... The defining feature of these experiences, in Jaspers' view, was an irreducible strangeness, and it was futile even to speculate about some central factor or deficit underlying the puzzling variety of schizophrenic symptoms.

It is odd that Jaspers, of all people, should have insisted so adamantly on the utter incomprehensibility of this condition; he, after all, was the most important champion of a *verstehende* psychiatry – one of interpretation or understanding.<sup>181</sup>

It is on the basis of such a critique that Jaspers' contribution to understanding the nature of 'true' delusions is ignored. In fact, just as the textbooks tend to employ his external criteria for 'delusion' but ignore the fact that he explicitly denies that such criteria in any way do justice to the psychological character of delusion, those who tend to note (correctly) that Jaspers considered delusions intrinsically 'ununderstandable' fail to notice that he spends some 10 pages (*GP* pp. 98-108) trying, precisely, to understand the character of primary delusions.

---

<sup>179</sup> Gregory Currie, *Imagination, Delusion and Hallucinations*, pp. 167–168.

<sup>180</sup> Richard Bentall, *From Cognitive Studies of Psychosis to Cognitive-Behaviour Therapy for Psychotic Symptoms*, p. 8.

<sup>181</sup> Louis Sass, *Madness and Modernism*, p. 17.

What this surely indicates is not that Jaspers has gone mad, but rather he is employing the concept of 'ununderstandable' *in a particular way*. For of course, on one reading of 'ununderstandable', *neurotic* symptoms such as compulsive handwashing, the anxiety found in depression, or phobic avoidance behaviour neatly fit the bill. It is in virtue of their being ununderstandable, *from a particular point of view*, that they are considered pathological. And neither is there any simple characterisation of this point of view; it will not do, for example, simply to say just that they are *rationally* inexplicable, for all of us, neurotic or not, are sometimes irrational.<sup>182</sup>

Turing to the examples given above the shallowness of the common reading becomes evident. With respect to the second quote, Jaspers admittedly does say that delusions cannot be '*sufficiently* understood in terms of the personality or the situation' (GP p. 107, my italics), but also argues that 'the criteria for delusion proper lie in the *primary experience of delusion* and in the *change of the personality*' (GP p. 106, italics in the original), or that '*Delusion proper* is incorrigible because of an *alteration in the personality*, the nature of which we are so far unable to describe, let alone formulate into a concept, though we are driven to make some such presupposition.' (GP p.105, italics in the original)<sup>183</sup>. With respect to the third quote, Jaspers' attempts to understand primary delusions are grounded both in his phenomenological approach, but also in his '*Verstehende*' psychology<sup>184</sup>, even if it is by the measured inapplicability of this method that delusions are ultimately judged. (Furthermore, Jaspers sometimes gave

---

<sup>182</sup> Much of ordinary mental life too is surely ununderstandable by reference to psychological criteria: for example, unless we subscribe to Freud's widely disparaged view of psychic determinism, most of us would be unwilling to concede that there are always psychological reasons why, when asked to pick a number at random, we pick the number we do, or why a certain jingle occurs to us at exactly the time it does, and so on.

<sup>183</sup> The tension between the first and later quotes can be resolved if we note first, the use of the word 'sufficiently' in the first quote, and note secondly that in the first quote Jaspers is most likely referring to the inability of primary delusions to be understood psychodynamically – as what the patient might be thought of as needing to believe given their fears, their self-image, their unconscious desires and so on – whilst in the later quotes he is using 'personality' in his sense as the 'totality of understandable connections', and it is in terms of a *change* in, and not some pre-existing *trait* of, the personality, that the delusion is to be understood.

<sup>184</sup> Chris Walker, *Delusion: What did Jaspers Really Say?*, p. 102.

reasoned criticism of those (Jung and Bleuler for example) who over-estimated the possibility of giving interpretative understanding of psychotic phenomena<sup>185</sup>.)

After supplying the unhelpful external criteria of delusion, Jaspers goes on to describe (GP p. 96):

two large groups of delusion according to their *origin*: one group *emerges understandably* from preceding affects, from shattering, mortifying, guilt-provoking or other such experiences, from false-perception or from the experience of derealisation in states of altered consciousness etc. The other group is for us *psychologically irreducible*; phenomenologically it is something final. We give the term '*delusion-like ideas*' [secondary delusions] to the first group; the latter we term '*delusions proper*' [primary delusions]

As well as being 'psychologically irreducible' or 'phenomenologically ...final', Jaspers considers primary delusions to be 'unmediated by thought' and 'ununderstandable'. In fact, all of these expressions are attempts to capture the same feature of true delusions, and to explain just what it is about these beliefs that is so peculiar and alien. Primary delusions are not understandable, i.e. are psychologically irreducible, because there is no possibility of grasping the manner in which they have been mediated by thought, and this because they just are not so mediated in the first place. They are not the product of mistaken reflection, they are not intelligible as beliefs which have been arrived at as an attempt to make sense of experience. (Secondary delusions or 'delusion-like ideas' are, by contrast, intelligible responses to pathological experiences (such as hallucinations) or to primary delusions or to extreme affects) – they are understandable when seen in the context of these experiences, feelings and beliefs.)

This view, then, contrasts with the position that (p. 96) 'all delusions are understandable in themselves and secondary'. Jaspers argues that this position leaves us (p. 97) 'still without an explanation of the essential nature of delusion ... we are offered only an understandable context for the emergence of

---

<sup>185</sup> As discussed below, the relevant issue is not in fact whether delusions can be understood given the preoccupations and motivations of the subject, but rather whether they can be understood as intelligible responses to perceptions or other beliefs. (cf GP p. 196, § (c)). More generally: Jaspers wanted to bring psychiatry back into the fold of the human sciences (concerned with *Verstehen* – understanding) whilst not excluding it from the natural sciences (concerned with causal and structural *explanation*); he felt that the latter sciences had somewhat taken over the discipline. (Today this would be viewed as a reaction against a purely 'medical model' of mental illness). But his main contribution to psychopathology stems from his *phenomenological* stance, a stance which, incidentally, is employed to such good effect by Sass himself.

certain *stubborn misconceptions* [my italics]. If these misconceptions turn into delusion, something new has to arrive'. In saying this, Jaspers is appealing to our intuition, for example, that a belief which is arrived at on the basis of an hallucination not recognised as such does not possess the requisite *strangeness* of a true delusion. If for example I (unawares) hallucinate my cat walking through the door, I will believe that my cat has entered the room. This however is just a wrong belief, and it is psychologically explicable (or 'reducible') in terms of the prior hallucination. It does not reflect a defect in my reason, in no way is it irrational, and it is in the circumstances fully understandable given my experience. It is not indicative of insanity, in contrast to delusion which is the *sine qua non* of psychosis.

A further example may help consolidate this sense of ununderstandability: I may view my neighbour's doorstep and see that the milk has not been taken in for a couple of days. I thereby conclude that he is either ill or away, and this conclusion can be stood in a rational relation to the evidence on the basis of which it was formed. When undergoing a psychotic episode, however, I may see the milk on my neighbour's doorstep and suddenly realise that I am the only survivor left on the planet after a nuclear holocaust (delusional perception). There is no way that this 'conclusion' can be rationally related to the 'evidence'; in this sense it is ununderstandable.

In arguing that delusions proper are ununderstandable, Jaspers is not saying that delusions cannot be understood psychodynamically. Thus (p. 196):

(c) We may well *understand from the context* how a delusional belief liberates an individual from something unbearable, seems to deliver him from reality and lends a peculiar satisfaction which may well be the ground for why it is so tenaciously held. But should we also try to make the actual formation of the delusion, as well as its content, understandable, any diagnosis of delusion becomes impossible, for what we have grasped in this case is ordinary human error, not delusion proper.

A delusion may in one sense at least then be considered within *motivational* psychology; believing that I am the king of Prussia or the world saviour may serve to promote my self-worth, which in the circumstances of my hospitalisation and social ostracisation may otherwise be none too great. But it would not be a delusion if the genesis of this belief could be rationally understood, if it were a rational

response to experience, to something that the pope once told me.<sup>186</sup> Even when a delusion can be psychodynamically understood, then, no implication follows that it can be made intelligible in terms of the standards of rationality: however sensible it might be for me to believe that I am the world saviour given the effect it has on my mood, nothing follows from this about the sensibleness of the belief *per se*, and it is in such negative terms that the delusion is to be assessed. The delusion does not make sense in terms of the personality, where by 'personality' is meant the totality of the understanding and beliefs of the subject; but it may make sense in terms of the personality, where by 'personality' is meant the dynamic factors imminent within character provoking a *need* to believe.

The inability of delusion to be understood as a rational response still leaves open the question of the nature of delusion. The psychodynamic hypothesis, for example, is just that – not a characterisation of delusion, of what it is that constitutes a delusion's bizarreness, but rather an hypothesis as to the rôle played by the delusion in the psychic economy of the subject. Contrary to what the above-quoted commentators suggest, Jaspers has a considerable amount more to say about the core (essential) psychological nature of delusion, although he frequently admits that (e.g. p.96) 'a clear presentation is hardly possible with so alien a happening'. The most essential criterion for delusions proper, he considers, is that they be grounded in *an altered experience of meaning* and in *a change in the totality of understandable connections of the personality*.<sup>187</sup> I shall consider these in detail in chapter 7.

### 3. Cognitive Theories of Delusion as Failures in Rationality

#### i. Empirical Theories and Evidence

An early yet typical proponent of the view that schizophrenic delusions are the result of defective reasoning processes was E. Von Domarus.<sup>188</sup> Von Domarus proposed that the relevant defect was in

---

<sup>186</sup> The difficulty in imagining circumstances which could make someone who was not the king of Prussia believe that they were shows how the psychiatric form/content distinction, at least as applied to *beliefs*, is not absolute.

<sup>187</sup> e.g. *GP* p. 106.

<sup>188</sup> *The Specific Laws of Logic in Schizophrenia* (1944). Referred to by Kemp et al., *Reasoning and Delusions* p. 398.

deductive reasoning with syllogisms, specifically, that schizophrenic patients were likely to make false assumptions about the common identity of two subjects on the basis of identical predicates.<sup>189</sup>

More recently, Phillipa Garety and David Hemsley proposed that some schizophrenic delusions are the product of biases in judgement, specifically in probabilistic reasoning. Their earlier (1988) theory proposed a specific deficit in Bayesian inference.<sup>190</sup> A typical procedure used to test such inference making would be the following. Subjects are shown two jars containing red and green beads; one jar has 15% red and 85% green beads, the other 85% red and 15% green beads. These are then removed from sight, and beads are removed from one of them one by one. Based on the colour of the beads removed the subjects are asked to guess from which jar they are removed; they can offer their judgement either when they feel certain, or they can offer an estimate of probability that one rather than the other jar has been chosen. Other more recent studies by Garety & Hemsley and colleagues have tested whether delusional subjects are more likely to jump to conclusions in reasoning, or to show biases in data gathering, than non-delusional controls.

A related study by Young & Bentall (1995) examined hypothesis testing in delusional subjects with visual discrimination tasks, requiring the subjects to successively narrow down hypotheses in response to feedback.

Even more recently (1997) Roisin Kemp and colleagues<sup>191</sup> tested delusional and non-delusional subjects on a variety of reasoning tasks. *Conditional reasoning* (using the form 'If p then q') was tested by asking subjects questions such as 'If she meets her friend, then she will go to a play. She does not meet her friend – What follows?'. Various manipulations of the questions enabled the tendency of subjects to make logical errors – such as affirming the consequent (She goes to the play, therefore she has met her friend) or denying the antecedent (She doesn't meet her friend therefore she doesn't go to the play) – to be evaluated. *Syllogistic reasoning* was tested with examples such as: 'People may be attracted to religion for many reasons including upbringing and life experiences. *No religious people are criminals.* Becoming a priest requires dedication but some may find the demands too much to live up to.

---

<sup>189</sup> E.g.: A poodle has hair on its head. I have hair on my head. Therefore: I am a poodle.

<sup>190</sup> Huq, Garety & Hemsley, *Probabilistic Judgements in Deluded and Non-Deluded Subjects*.

<sup>191</sup> Op cit.

*Some priests are criminals.* Let us agree that what has just been said is true. Does it follow that: *Some priests are not religious people?* Probabilistic reasoning was tested by presenting scenarios, and asking subjects to evaluate the likelihood of several options, such as: 'Sally is 29 years old. She ran away from home at the age of 15 because she got pregnant. She is sexually attractive and has had many partners. Recently she has lost a lot of weight and has had to go into hospital for tests. Sally ... a) is a famous high court judge; b) is a teacher in a primary school; c) is a teacher in a primary school and has AIDS.'

Various other studies in the experimental paradigm have been carried out on the reasoning of delusional subjects; these can be found summarised in a recent paper by Garety and Daniel Freeman.<sup>192</sup>

The results of such studies are not uniform. Von Domarus' hypothesis of a failure in syllogistic reasoning has for example been called into question by a study by Brendan Maher<sup>193</sup>, who argued that non-delusional subjects matched for education and IQ were as likely as delusional subjects to make logical errors. This finding was essentially replicated by Kemp et al's tests of syllogistic reasoning; although they found that delusional subjects were more likely to endorse fallacies in conditional reasoning, their overall conclusion was (p. 402) that 'all subjects [i.e. including controls] displayed considerable irrationality and a propensity to go with prior beliefs rather than reasoning through a problem', and that (p. 398) 'Differences in reasoning between deluded patients and controls are surprisingly small'.

Garety et al.'s studies of probabilistic reasoning with the coloured beads appeared to reveal a propensity towards faulty inference by the delusional subjects, who were far more ready to form probability judgements after very few (one or two) beads had been examined than were non-deluded controls. Nevertheless, Maher has pointed out that Garety et al.'s delusional subjects were actually *more* rational than non-delusional controls (!) in making a judgement about the jar of origin for the coloured beads after only two beads had been removed. For after two beads with the same colour have been drawn, the Bayesian probability that the beads were drawn from the jar with 85% beads of that colour is 97%. (Non-delusional subjects tended to wait for more corroborating evidence than was reasonably required).

---

<sup>192</sup> *Cognitive Approaches to Delusions: A Critical Review of Theories and Evidence.*

<sup>193</sup> *Delusions: Contemporary Etiological Hypotheses.*

According to Garety et al.'s (2000) review paper, studies of delusional subjects tend to reveal that whilst there is little evidence of a general probabilistic reasoning bias, delusional subjects do tend to accept hypotheses on the basis of less evidence than members of the general populace – do, that is, tend to jump to conclusions.

Before conclusions are drawn, however, as to the rôle of deficient reasoning in the formation and maintenance of delusions, it should surely be remembered that most of the delusional patients in these studies suffered from *schizophrenia*. This is to say that they suffer from a variety of psychological problems such as distractibility, anxiety, or 'capture' by irrelevant environmental stimuli, and their relatively poor reasoning in such tasks, which in any case is not that different from controls<sup>194</sup>, may have little or nothing to do with their delusions. Furthermore, the delusional subjects in Garety et al.'s trials were also far more likely to abandon their 'rashly' adopted hypotheses and form new ones in the light of contradictory evidence than were controls – something which would not be expected if such reasoning processes were at work in the formation of delusions, which are notoriously hard to shift. A reasonable consensus opinion then, concerning the rôle of reasoning deficits in delusions, on the basis of the experimental findings in the literature, is that such studies have (at best) found *associations* between delusion and poor reasoning, and not evidence of a causal rôle.

## ii. Critique of Suppressed Premises

At this point I should like to step back from the experimental paradigm and restate one of the questions with which I began. Which is: Are those beliefs that have the specific quality of bizarreness that qualifies them as delusions, and especially those delusions that are found in the schizophrenic population, the kinds of beliefs that are *plausibly* explained by general reasoning deficits? Consider some typical schizophrenic delusions:

It suddenly occurred to me one night, quite naturally, self-evidentially but insistently, that Miss L. was probably the cause of all the terrible things which I have had to go through these last few years (telepathic

---

<sup>194</sup> Kemp et al, *Reasoning and Delusions*.



influences, etc.). I can't of course stand by all that I have written here, but if you examine it fairly you will see there is very little reflection about it! Rather everything thrust itself on me, suddenly, and totally unexpectedly, though quite naturally. I felt as if scales had fallen from my eyes and I saw why life had been precisely as it was through these last years.<sup>195</sup>

A woman said, 'every night blood is being injected out of my arms' (*sic*). When asked for her evidence she explained that she had little brown spots on her arms and therefore knew that she was being injected. The interviewer looked at the spots on her arms, rolled up his sleeve and showed her spots identical in appearance on his own arm. He said that they had been on his arm as long as he could remember and were called 'freckles'. She agreed that both sets of spots looked similar and accepted his explanation of his own spots, but still insisted that her freckles proved that she was being injected in her sleep.<sup>196</sup>

SW ... expressed the beliefs that television and radio referred to him and certain records on radio were chosen deliberately to remind him of his past life. He was convinced his food was poisoned and felt his head and genitals were being compressed as a result of an aeroplane flying overhead.<sup>197</sup>

At the police station I had the impression that I wasn't at the station but in the Other World; one official looked like death himself. I thought he was dead and had to write on his typewriter until he expiated his sins. Every time the bell rang I believed they were fetching away someone whose lifetime had ended. (Later I realised the ringing came from the typewriter as it reached the end of the line.) A young policeman had a pistol in his hand; I was afraid he wanted to kill me. I refused to drink the tea they brought me as I thought it was poisoned. I was waiting and longing to die ... it was on a stage, and marionettes are not human. I thought they were mere empty skins ... the typewriter seemed upside down; there were no letters on it, only signs which I thought came from the Other World.<sup>198</sup>

Considering such examples, two arguments suggest themselves: Firstly, the deficits in rationality manifest in delusions hardly seem to be *general* deficits, but rather really quite *specific and localised*.<sup>199</sup> Secondly, the irrationality of the delusions hardly seems to be the product of a *failure in reasoning*,

---

<sup>195</sup> Example from Jaspers' *General Psychopathology*, p. 103, in the section on delusional ideas.

<sup>196</sup> From Andrew Sims' *Symptoms in the Mind*, pp. 107–8, example of a delusional percept.

<sup>197</sup> From Frith's *Cognitive Neuropsychology of Schizophrenia*, p. 2.

<sup>198</sup> Jaspers pp. 101–2, on delusional reference.

<sup>199</sup> This especially so with *encapsulated* delusions.

*inference*, or a tendency to *jump to conclusions* – at least, not if these terms refer to a disruption of normal procedural reasoning. I shall consider these in turn.

Striking about schizophrenic delusions is that they tend to be restricted to the interpersonal domain. This is clearly the case for paranoid delusions, delusions of reference etc, and it may not be an exaggeration to say that all true delusions are self-referential in character.<sup>200</sup> For this reason it seems implausible to explain the formation of delusions by a general failure of inductive reasoning. Furthermore, it would not do for the deficit-in-reasoning theorist to suggest that delusions are the product of the product of a failure in specifically social reasoning. For even if such failures in our capacities to understand one another as subjects and agents – in our ‘theory of mind’<sup>201</sup> – obtain in conditions such as schizophrenia, they could not account for the *specificity* of the delusions in question. Delusions tend not to be found throughout the whole of the interpersonal sphere, and are rather restricted to particular preoccupations; furthermore and in any case, general social reasoning may not be impaired.<sup>202</sup>

Let us now consider the question of the *a priori* appropriateness of an explanation of delusion in terms of reasoning deficits: If we return to 1913 we find the following on page 97 of Jaspers’ *General Psychopathology*, where he is discussing the theory that delusions are a product of a weakened intelligence:

We always tend to look for the logical errors and blunders in the paranoid patient’s thought in order to prove some such weakness. Sandberg, however, pointed out quite rightly that paranoiacs have by no means a poorer intelligence quotient than healthy persons and in any case the mentally ill person surely has as much right to be illogical as the healthy one. It is wrong to consider the failure in reasoning a morbid symptom in one case but

---

<sup>200</sup> Whilst overvalued and delusion-like ideas (‘secondary’ delusions) may be less self-referential. (See Sims, *Op. Cit.* p. 128.)

<sup>201</sup> Frith, for example, suggests that several schizophrenic symptoms could be traced to a failure in ‘theory of mind’ in schizophrenia, and to this extent compares the condition with childhood autism, accounting for the differences between the conditions by viewing schizophrenic symptoms (delusions of reference etc.) as due to attempts to use a once healthy but now disrupted theory of mind module, and the autistic symptoms (social withdrawal but no paranoid delusions) as products of a failure to develop ‘theory of mind’ skills in the first place. (Frith, *Cognitive Neuropsychology...*, pp. 117–126.)

<sup>202</sup> For similar objections to accounts of delusion formation in terms of general deficits in procedural rationality along the lines given above, see I. Gold and J. Hohwy, *Rationality and Schizophrenic Delusion*, section 5.2.

normal in the other. Actually we find every degree of mental defect without delusions of any kind and the most fantastic and incredible delusions in the case of people of superior intelligence. The critical faculty is not obliterated but *put into the service of the delusion*. ... With delusion proper there is material falsification while formal thinking remains intact. Where there is formal thought disturbance, then misapprehensions, confused associations and (in acute conditions) the wildest notions may follow, which as such do not have the character of delusion proper.

More recently (1995), Sims wrote, on p. 142 of *Symptoms in the Mind*:

A schizophrenic delusion is not a simple defect of reasoning... It is an assumption about the world the patient inhabits, which he does not create by a process of logical conscious thought from premises distorted by emotion. The starting points of his thinking are already 'deluded' and his logic elaborates from this basis.

It has long been a contention, then, of the most respected of psychopathologists that the reasoning of the delusional patient is not generally defective – that it is even 'put into the service of the delusion' – and that the core of the delusion is to be found in the 'starting points' of this reasoning.

Consider too the claim that delusions may result from a tendency of the schizophrenic subject to 'jump to conclusions', and consider this in the context of Garety et al.'s above-mentioned experiment with the coloured beads. Garety's delusional subjects were somewhat more prone to make *precipitate* judgements, to come to conclusions on the basis of inadequate evidence – inadequate in the sense of *too little*. But whilst some 'secondary' paranoid delusions (paranoid delusion-like ideas) may be accurately described in terms of jumping to precipitate conclusions, true schizophrenic delusions seem to be of a quite different form. True delusions tend to have the form: I saw the traffic lights change and just knew that I was the next world saviour, or: 'I knew that my wife was unfaithful immediately I saw the bulb had gone out.'<sup>203</sup> To describe such thoughts as the product of 'jumping to conclusions' cannot but sound somewhat ridiculous, for it is not at all easy to understand the antecedent *as* a conclusion. I am not simply *precipitate* in assuming that my brain has been replaced by a swarm of bees, *even if* it seems to me that I hear certain sounds and feel certain sensations within my scalp. We do not have to do with too *little* evidence, but with a *failure* of what may be produced by way of evidence (change of the traffic

---

<sup>203</sup> Example from Sims, *Op. Cit.* p. 142.

lights) to stand in a sensible evidential relation to that (my being the next world saviour) for which it is adduced.

The principle difficulty with cognitive theories of delusional irrationality then is not that, even if delusional patients tend to make mistakes in their reasoning, there is no evidence that these are responsible for their delusions. The problem is rather that we should not really know how to understand a true delusion as the consequence of reasoning, faulty or otherwise. If a belief is viewed as a reasoned response or conclusion (whether well or poorly reasoned) it is hard to see how it could also be a delusion.

#### **4. Delusion as Rational Response**

##### **i. Maher's Theory of Delusions as Normal Theories**

If faulty reasoning doesn't and/or cannot explain schizophrenic delusions, an obvious alternative to investigate is the view that delusions constitute reasonable interpretations of *abnormal experiences* of either bodily (proprioceptive) or worldly (perceptual) form. (A related position has already been examined in the last chapter, where the theory of Frith and G&S that delusions of thought insertion and of hearing voices are consequent upon a reasonable interpretation of an *abnormal experience of one's own mindedness*.) This is clearly a natural suggestion, because many delusional subjects, especially those with schizophrenia, suffer a variety of unusual experiences (hallucinations, experiences of depersonalisation, derealisation, passivity experiences etc.), and it has long been recognised that many delusions, or at least '*secondary delusions*' or '*delusion-like ideas*', are readily intelligible as reactions to strange experiences. Jaspers, for example, writes of that group of (secondary) delusions that '*emerges understandably from preceding affects, from shattering, mortifying, guilt-provoking or other such experiences, from false-perception or from the experience of derealisation in states of altered consciousness etc.*'<sup>204</sup> On the other hand there clearly exist delusional conditions that do not involve

---

<sup>204</sup> *General Psychopathology*, p. 96; c.f. also Kraepelin's views on delusion in his *Dementia Praecox*.

abnormal experiences<sup>205</sup>, so the model can hardly have universal application. What needs to be investigated is whether those delusions that are most characteristic of schizophrenia are plausibly to be understood as reactions to unusual experiences.

In the cognitive literature the experiential theory is most closely associated with Brendan Maher.<sup>206</sup> Maher proposes that delusions are best understood as 'normal theories' developed to account for puzzling experiences. Such experiences will generally be puzzling simply because they are so unusual, but Maher also suggests that in some cases psychotic subjects may be disposed – for *neurological* reasons – toward finding perfectly normal experiences a source of bemusement. This is to say that a psychotic subject may sometimes find some experience puzzling for *no* (folk-psychologically specifiable) *reason* at all; their puzzlement at, or their sense of the significance of, some perfectly normal experience being simply a function of a failure in the matching of (p. 21) 'one neurally defined template (the expected sequence of observations) with another neurally defined template (the experienced sequence of observations).' Nevertheless, according to the theory this possibility relates only to certain 'coincidence delusions'; otherwise delusions are best understood as responses to genuinely unusual experiences.

It would be fair to say that Maher does not surround his experiential theory with demonstrations of the experiential grounding of delusion. The argument goes, rather, from the lack of grounds for the contrary hypothesis that delusional thought is in itself 'aberrant' to the conclusion that delusions are (p. 20) 'normal theories'. In what follows I shall (predominantly) not question the view that delusions may have an experiential ground, but rather query whether Maher's grounds for supposing that delusional thinking is not in itself aberrant are adequate. In particular, Maher seems to identify the view that delusional thought is *aberrant or irrational* with the idea that delusions are the product of *faulty reasoning* – in other words, with the view outlined in section 2 above.

This is most striking in the first of the 'formal propositions of the [delusions as normal theories] model':

---

<sup>205</sup> Chapman and Chapman 1988. The most obvious example of delusions that are clearly not based in abnormal experience are *delusional perceptions* – which by definition are grounded in *normal* experience.

<sup>206</sup> Maher (1974) *Delusional thinking and perceptual disorder*. Also (1988) *Anomalous Experience and Delusional Thinking: The Logic of Explanations*. Page references in the text are to the latter publication.

1. Delusional thinking is not, in itself, aberrant. This means that the cognitive processes whereby delusions are formed differ in no important respect from those by which nondelusional beliefs are formed.

This clearly identifies the question of the *aberrance of delusional thought* with the question of the *unruliness or propriety of the cognitive processes* by which delusions are formed. The rest of the model outlines the view that delusions are normal theories, and considers the process by which such theories are developed; to summarise with extracts:

2. Delusions are best thought of as theories .... that serve the purpose of providing order and meaning for empirical data obtained by observation.

3. The necessity for a theory arises whenever nature presents us with a puzzle. Puzzles arise when a familiar and hence predictable sequence of observation fails to occur in the expected fashion. ...Puzzles are surprises. ... when there is a discrepancy between what we expect to observe and what we do observe, we experience the discrepancy as significant. We notice it; we are brought into a state of alertness and tension; it puts us into what might colloquially be termed a "search mode."

4. Puzzles demand explanation ....

5. When an explanation for such a puzzle has been developed, it is accompanied by marked feelings of relief and tension reduction...

6. Data obtained subsequently that contradict the explanation create cognitive dissonance and are unwelcome. Data that are consistent with the explanation reduce dissonance and are given particular status in the explanation.

7. Theories will be judged delusional by others if (1) the data upon which they are based are not available to those who are judging ... [or] (2) the data are available but most observers do not experience puzzlement or sense the significance that the patient does.

10. A delusional theory, like other theories, is not readily abandoned until it can be replaced by a theory that better explains the experiences that the patient is having. Hence the folk-clinical observation that delusional patients do not readily abandon their theory in the face of critical contradictory evidence does not indicate a pathology of reasoning.

## ii. Critique of Maher's Theory

In his 1988 paper Maher considers various standard criticisms of his model and offers rebuttals. One objection which he considers is that a demonstrable *correlation* between delusion and abnormal experience does not in itself offer proof of a causative influence (on the delusion by the hallucination). Maher concedes the objection, making no rebuttal. This however seems a precipitate capitulation, for it is surely the case that many delusions (or at least 'delusion-like ideas') are at least manifestly *intelligible* given the subject's abnormal experience. A problem here is that the cognitive model seems to sit between two forms of explanation: on the one hand it employs the standard (scientific) rhetoric of 'causal influences' and (p. 25) 'etiological inferences', but on the other hand it is only natural to construe and understand the issues in a *folk*-psychological idiom – to see, in other words, whether an appeal to abnormal experiences can make the delusion *intelligible*. Maher urges 'the necessity of establishing the nature of the pre-delusional experience if we are to make etiological inferences', which indicates that the 'causal' issue is to be tackled in terms of *temporal priority*. Nevertheless it is hard to imagine how this temporal ordering could be reliably established; furthermore it is unclear that talk of a temporal priority – and of a 'pre-delusional' experience is always apt, for it may be that the delusion and the experience either arise together (in which case it may still be that the delusion is *intelligible* in terms of the experience), or perhaps that the experience and the delusion are *identical*.

A second objection, to which Maher offers a rebuttal, asks why the delusional patient is prone to reject what is surely the more natural interpretation of their experience, namely that provided by the psychiatrist (that the patient has suffered an hallucination). Given that they resist such an interpretation, and given that this resistance is characteristically cited in textbook definitions of delusion (delusions are 'unshakeable', 'impervious', held with 'extraordinary conviction' etc.), it is surely unreasonable to suppose that delusions are to be understood simply as rational (or at least typical) responses to unusual experiences. Against this objection Maher partly capitulates, and offers up the rest of his model: that delusional patients may be prone to find coincidences more striking, puzzling and in need of explanation than the 'normal' population. But the main of his response consists in noting that people *without* delusions (including scientists) are frequently just as obstinate and unwilling to give up their ideas once they have taken shape; that a delusion may in fact be a very good, and for the subject a very natural,

explanation of the experiences in question; and that non-deluded people (again including scientists) frequently prefer superstitious to scientific explanations of strange events.

Along with the important observation of the previous section, that the general reasoning of the deluded patient tends not to be abnormal, and not therefore a helpful locus of explanation for the formation of delusions, the last-mentioned empirical counter-objection can be readily accepted. What is doubtful, however, is whether this really allows the case to rest against the view that delusions are the kinds of beliefs that in the circumstances are perfectly reasonable. For from a perfectly natural pre-theoretical point of view, delusions are clearly *extremely bizarre*. As Gold and Hohwy put it<sup>207</sup>, with no intention of prejudicing or prejudging the issues: 'That delusions *are* irrational we take to be obvious. Like ... a desire for a saucer of mud [Elizabeth Anscombe's example in *Intention*, section 37], ... delusions ... are *paradigms* of irrationality.' The question that needs to be asked, then, is whether the criteria of unreasonableness, irrationality or aberrance that are employed within the cognitive theory, and against which the rationality of delusional beliefs are assessed, are themselves adequate.

As noted above, one of the arguments Maher gives for supposing that non-deluded people can be just as irrational as deluded people is that people with non-delusional (yet false) beliefs can be just as reluctant as delusional people to give up the beliefs in question. The use of intransigence as a criterion for irrationality is both fairly intuitive (maintaining a belief in the face of vast amounts of countervailing evidence just *is* irrational) and is also drawn from the common textbook and DSM-style definitions of delusion. The relevant question however is whether intransigence is a relevant criterion for the type of irrationality one meets in delusional subjects.

Recall again Jaspers' distinction between the internal and the external characteristics of delusions. Various characteristics (such as falsity, cultural incongruity etc.) may well be typical features of delusions extensionally characterised, but because they do not necessarily obtain, and because intuitively they fail to capture the essence of delusion, they count as external and not internal marks. May not the same be said of the stubbornness with which delusions are held, their imperviousness to reason? Perhaps, in other words, a patient is judged delusional not simply because they have stubborn impervious beliefs – something of which we may all be guilty – but because they have (stubborn, impervious) beliefs *like that*.

---

<sup>207</sup> *Rationality and Schizophrenic Delusion*, p. 149.



Namely: *delusions*, the *internal* characteristics of which are yet to be determined. And perhaps what both a psychiatrist and a layperson find incredible in the delusional subject's intractability is not that they simply hold strongly to their beliefs, but that anyone could, except for a fleeting moment on waking, or whilst under the influence of a psychedelic substance, or during the aura of an epileptic attack, maintain with any degree of seriousness a conviction in the *kind* of strange beliefs to which the delusional subject subscribes. Just because, then, we are all guilty to some degree to a perverse stubbornness in giving up our beliefs does not mean that a delusional patient should be judged perfectly sane in cleaving in the same way to theirs.

The same quality of strangeness that characterises delusions offered as explanations for bizarre experiences also characterises delusions that Maher explains in terms of a hyper-vigilance on the part of the delusional subject to the occurrence of coincidences. Two clinical examples given by Maher (pp. 29–30) are:

1. A patient who, finding himself in front of a house numbered 11 on Armistice Day, November 11<sup>th</sup>, is struck by the coincidence and concludes that he was responsible for World War 1.
4. A patient who, noticing that the letters of his name, when rearranged as an anagram, produce the name of a famous American Indian chief, concludes that he is the reincarnation of that man.

But whilst these examples may well have a source in hypervigilance, the problem remains that Maher provides no explanation of the patent strangeness of the delusional 'conclusion' in relation to the 'premises'. Thus anyone may be struck by the coincidence of the number of the house they are visiting and the (significance of the) day of the month, but it is hard to understand why this should lead someone to suppose that they are therefore responsible for the first world war. Indeed it is hard to understand this conclusion *as* a conclusion given the nature of the premises. The same goes for those characteristic delusions of schizophrenia found in delusional perception: I see a dog lift up its leg and straight away understand that the world is going to end (Schneider). Not only does the very existence of delusional perceptions represent a negation of Maher's theory (delusional perceptions are by definition delusional interpretations of *perfectly normal* experiences), but the 'and' in Schneider's patient's report hardly relates a conclusion to a premise in the way in which a piece of reasoning might.

Consider some of the other propositions of Maher's model. Proposition 7 argues that certain theories will be diagnosed as delusional either if the data on which the theories are based are unavailable to others, or if the data are available but others do not sense the significance of the data that the delusional patient does. Now a psychiatrist may diagnose a patient as delusional because they believe that they are the reincarnation of a famous American Indian chief. To do so simply for this reason would however be precipitate, for many people believe themselves to be reincarnations of previously existing persons, and are not to be judged delusional on this basis. What is relevant for the diagnosis is not so much the content of the delusion, but rather the *rational structure surrounding it*. Thus I may believe myself to be a reincarnated native American chief because I have what seem to me to be memories of rituals which I have not experienced in my life time, or because I seem to have knowledge regarding the preparation of herbs or the structure of Amerindian societies that I do not recall having gleaned from books or the television. Contrast the delusional subject who comes to this conclusion on the basis of the letters of their name; what is hard to understand here is how the premises relate to the conclusion. To be sure, it is not that the patient is to be judged delusional because their reasoning – construed as inference making – is faulty. Rather, it is hard to understand why what is proffered as a reason should *be* a reason, hard to understand what is said as a case of reasoning at all.

Returning to the first of Maher's propositions; whilst it may well be true that the 'cognitive processes whereby delusions are formed differ in no important respect from those by which nondelusional beliefs are formed', it is hard to understand this as a gloss for 'Delusional thinking is not, in itself, aberrant'. Whilst what is genuinely to be counted as the reasoning of the psychotic subject may be perfectly intact, it is hard to understand how, for example, the realisation that my wife is cheating on me consequent upon the light's being turned on or off could count as a piece of reasoning. It is not that a mistake has been made, nor that a logical blunder has been committed; nevertheless, we are left with a paradigmatic case of an irrational belief. Even if the patient had noticed a pattern in the turning on and off of lights, unless this can rationally be related to the cheating behaviour of the spouse (perhaps the light switchings spell out a message to a love over the street in morse code, perhaps they indicate that the spouse is returning from a secret liaison in the middle of the night and has to turn the lights on), the step from premise to conclusion fails to count as a piece of reasoning. It is the patent strangeness of such

beliefs, along with the psychotic lack of insight indicative of a failure to correct *this* strange sort of belief, that prompts the delusion diagnosis.

Ultimately, then, Maher's account fails as a theory of delusion not because it fails as a potential explanation of how *some* strange beliefs might be formed, but because it fails to account for the genesis of *those* beliefs that have the far deeper strangeness of delusions. Delusions might (proposition 2) be *offered* as 'theories' that provide 'order and meaning for empirical data', but in the final analysis it is just too hard to understand them *as* theories.

## 5. Conclusion

Cognitive theories of delusion tend to give the impression that there are only two alternatives to be considered. Delusions are to be seen either as rational responses to abnormal experience or as the product of defective reasoning. This perhaps is not surprising given the cognitivist's philosophy of mind which seems in some way to inform the cognitive psychologist's psychological theories. That cognitivist conception of our cognitive engagement with the world is made up of two components: a passive perceptual encounter with the world (perceptual 'input'), and a frequently subconscious processing of such inputs (in 'cognition') according to rules of reasoning – either formal (algorithms) or informal (rules of thumb, 'mental models'). If the formation of those irrational and strange beliefs that are delusions is to be inquired into in this theoretical context, there are only two choices: delusions are the product of defective reasoning procedures or defective input.

As we have seen, however, delusions cannot be theorised in either of these ways. They cannot be seen as rational responses to abnormal experiences, because (amongst other reasons) this would not explain their patent *irrationality*; at best such a form of explanation could address the formation of secondary delusions, or 'delusion-like ideas'. And they cannot be seen as failures of inferential reasoning, because the reasoning of delusional subjects is generally unimpaired, because an intact faculty of reason may actually be put into the service of delusion and aid in its elaboration, and most importantly because a characterisation in terms of a failure in reasoning – such as jumping to conclusions – both fails

to capture the *unmediated* character of the primary delusion, and once again makes the delusion *all too intelligible*, makes it hard to understand such a belief as possessing the strangeness of a true delusion.

Part of the source of the troubles with the cognitive theories comes from their implicit take on the psychopathological issues, supposing as they seem to that all delusions are secondary or that they can be reliably identified by their external characteristics. If true schizophrenic delusions could be so identified – as false, unusual and stubbornly held beliefs which could arise any old how – then the cognitive theories would at least have a chance of success. But another part of the problem with the psychological theories comes from their implicit take on certain philosophical issues, in as much as they seem committed to the cognitivist's (explicit) view that our i) epistemic and ii) rational grounding in reality is a matter of our i) correctly *representing* the world and ii) our employment of such representations in correct *reasoning*.

The latter (ii) suggestion has already been criticised in this chapter; the suggestion has been that the paradigmatically irrational character of true delusions cannot be realistically explained as a product of faulty *reasoning*. The following chapter (7), being devoted to an account of schizophrenic disengagement, tackles the earlier suggestion. Here the general argument from the first half of the thesis – that representational knowledge presupposes a background of praxical know-how – will be re-deployed. The cognitivist of course supposes otherwise, and, starting with a disengaged conception of human understanding, attempts to provide a theoretical account of mental representation and intentional action. Our everyday understanding becomes underpinned by propositions embedded in a folk-theory. (Maher's view that 'delusions are best thought of as theories' can be understood in this context.) Against the cognitivist's 'theory-theory' (their theory that our everyday understanding embodies a theory) it was argued that the perceived need for a theoretical account was only a function of the implausible and disengaged conception of understanding being provided, and that, since theory will always presuppose praxis, theory cannot be used in its explication.

The argument however does not all need to come from the side of philosophy. For one of the morals of the critique of the cognitive theories is that their failure to understand delusion reveals that a cognitive theory cannot reach the *depths* of the schizophrenic condition. The theorisation offered leaves us confronting an individual who is *not* really out of touch with reality, does not suffer the schizophrenic's

fundamental alienation from the world, from others and from themselves, but has really just made a few bad mistakes, or who has not reasoned things through correctly. The account of rationality and epistemology developed in the first half of the dissertation stressed that fundamental belief and understanding presuppose not internal representation or theory, but rather an *integrity of the Background*, and the praxical engagement of the subject with the world. What remains to be seen, then, is whether the psychotic core can be more adequately theorised in such terms.

## **Ch. 7. The Grammar of the Divided Self**

### **1. Introduction: Schizophrenia and the Fragmentation of the Background**

The critique of cognitive theories of schizophrenia is now at an end, and in this sense the brief of this dissertation has been met. Along the way however various unanswered questions have been raised about the nature of schizophrenic experience. In what follows I hope to show that the theoretical framework deployed in the critique of cognitivism contains the resources required to answer these questions and to provide a general theory of the core of psychosis in general and of schizophrenia in particular.

It is important to affirm at this point that, unless they are couched in the most general form possible, the questions that still remain to be answered are not the same as those that occupied the cognitive theorist. Parts 1 & 2 of the dissertation stressed that many epistemological and psychological theories of mind were a product of setting off on the wrong theoretical foot. A certain sort of understanding – of how we think, see, recognise, act – was sought, but the perceived need for this understanding was only generated within an alienated conception of mind. It is not just the answers, then, that were found wanting, but also the questions.

Part 3 introduced this philosophical critique into a consideration of psychological theories of schizophrenia. Here it was argued that an alienated conception of mind, sometimes in conjunction with an inadequate phenomenology, made for an appearance of an understanding of schizophrenia that was in fact illusory. To rehearse some of the main themes:

Disorders of intentional action, thought (in formal thought disorder) and emotion (in for example incongruity of affect) were theorised by the cognitive scientist as disorders in the expression of thought, emotion or plans. An alienated conception of mind encouraged the view that the intentional states of a subject are internal states in the sense that they are to be individuated in isolation from the activity, action and expression of such a subject. Given this conception it was thought possible to encompass the subject with schizophrenia within our normal psychology: disorders of affect and thought were reconceptualised as disorders of ‘output’ – in the mechanics of the expression of ‘inner’ states in ‘outer’ behaviours. This is one way of trying to spell out what I shall call the ‘schizophrenia intuition’ – the idea that the mind of

the subject suffering from schizophrenia is in some way *split*. The cognitivist's dualism of inner and outer cannot however be sustained. Intentions and feelings cannot be relegated to a self-contained inner world; their behavioural expression is an integral part of their identity. Because of this the schizophrenic impairments cannot be theorised as merely expressive; the schizophrenia intuition cannot be cashed out as a dissociation between thought and language or between emotion and facies. The split rather strikes at the heart of the self, at the integrity of thought and affect themselves; the schizophrenic condition cannot therefore be assimilated into normal psychology.

Hallucination and thought insertion were cognitively theorised as failures in inner sense. Once again the aim of the cognitive psychologist is to make the symptoms intelligible as dysfunctions in normal mechanisms, mechanisms which this time allow for a form of self-knowledge. In this way the condition can be assimilated into normal psychology: we can relate to the schizophrenic experience as something that can easily be understood to happen if our 'self-monitoring mechanism' were to become impaired. That is to say, the question 'Why does the subject with schizophrenia believe that thoughts are coming into their minds which are not their own?' is given a straightforward answer. Unfortunately the answer provided only makes sense within an alienated psychology (placing the mind not merely inside the body, but this time the subject behind their own mind looking in) which denies us the inalienable authority we enjoy in self-ascriptions simply by virtue of being true subjects.

Similarly with delusion: delusions were theorised by the cognitive psychologist as certain rather drastic mistakes. Once again the theoretical framework aimed to make delusions psychologically intelligible – as the kind of thing a non-psychotic subject might think if they too had the abnormal experiences of the patient with schizophrenia, or if they too became rather hasty in forming conclusions. In this way the delusional condition again becomes assimilated into normal psychology. However the phenomenology of delusion reveals it to resist this theorisation: delusions are simply too strange to be accommodated in this way. Furthermore, once it has been accepted that delusions represent *deep* failures in the delusional subject's contact with reality, a philosophical reason for the failure of the cognitive theories surfaces. The cognitivist's disengaged conception of the subject theorises contact with reality as based in theory, hypothesis, representational knowledge and representational belief. These however

cannot be the ground of our reality contact, for they presuppose (and so cannot explain) our worldly orientation.

The questions which the cognitive theorist asks and answers concerning schizophrenia, then, are of the form 'Why does the patient think that X?' It is just these questions that, if the arguments so far have been correct, cannot be answered. Nevertheless the following questions still remain: If the schizophrenia intuition is not to be spelled out in terms of the dissociation of faculties or of inner from outer, in what do the internal splits in the schizophrenic mind consist? If delusion is not a matter of false belief, then what is it to be deluded? In what does the delusional subject's lack of contact with reality consist? If we are not to understand or make sense of the patient's reports of thought insertion without making such reports themselves make sense, if we are not to view them as descriptions of understandable experiences, then how are they to be understood? In all such cases the understanding sought is not to be provided by a causal theory or explanation; what is rather needed is a clarification of what it means to suffer thought insertion, hallucination, delusion and thought disorder.

In Parts 1 & 2 of this dissertation the suggestion was made that expressions and actions are not to be considered the manifestations of intention, feeling, knowledge, understanding etc. because they are the causal products of inner states called intentions, feelings etc., but because they are situated in a pervasive yet unassuming Background context. This Background provides the context within which ascriptions of propositional attitudes can be made; at the highest levels it consists of other such attitudes, at middle levels it involves the know-how manifest in complex action and language use, at the lowest levels it is manifest in a necessary pre-intentional coherence of movement, reaction and disposition. Part 3 found that the view of mind as a realm of inner states was unhelpful for understanding schizophrenia. In this part (4) I shall suggest that psychotic phenomena need to be understood as *internal ruptures in the Background itself*.

Two things tend to stand in the way of such an analysis and in what follows I shall be on my guard against both. On the one hand it is hard to disengage our normal and natural psychological mode of understanding which tacitly (yet perfectly licitly) imports the Background as a frame of reference within which psychological questions can be assessed. The Background is not itself easily pulled into view, nor is the stance which uses it as a tacit backdrop easily set aside, as it must be when it itself is the object of



investigation. (It is because the Background is so pervasive yet tacit that, it was argued, the cognitivist is unaware that their theorisation of propositional attitudes etc. as stand-alone inner states cannot succeed. The propositional attitudes might *theoretically* be taken as self-contained states, but this is only because we naturally, inexorably and tacitly bring our *pre-theoretical attitudes* to content-bearing states toward them, thereby disguising the inadequacy of the theoretical treatment.) On the other hand the psychopathological phenomena (delusions, passivity experiences) themselves invite certain characterisations that whilst perhaps somewhat paradoxical at least make use of folk-psychological concepts and viewpoints.

Cognitive theorisations tend to attempt their psychological explanations by making the paradoxes of psychosis go away; this however is a mistake. The patient who hears voices commenting on their thoughts and actions is talking to themselves without realising it; on the one hand this is true by definition, but on the other hand it is paradoxical since when we are dealing with what is said *in foro interno* there is (contra the cognitivist) no longer a distinction between speaking and listening which makes for the possibility of self-misunderstanding. The patient with thought insertion says that they have thoughts in their minds which are not their own. On the one hand this is just what it is to suffer the symptom, but on the other there is no such thing (contra the cognitivist) as experiencing or not experiencing the self-generation of thoughts. The patient with delusion believes that they have an army of Russian soldiers sitting inside their abdomen. But what can it mean to believe something this *outré* – aren't our beliefs governed by a constitutive principle of rationality? Another patient both believes and does not believe that they are the king of France; or they are both in love with and absolutely deprecating of their fiancée. But what does it mean to believe or feel such totally irreconcilable beliefs or feelings?

What I believe is important in coming to understand psychosis and schizophrenia in particular is not to try and make these paradoxes disappear by making the phenomena cognitively intelligible, but rather to come to a clarificatory understanding which maintains the paradoxes and reveals how and why they come about and why they are *constitutive* of the phenomena in question. Despite the difficulties involved in bringing the Background into focus, the theory will be that viewing psychosis as a fragmentation of the Background itself allows us to *come to terms* with the paradoxes and with the psychopathological

phenomena and in this sense provides psychological understanding without solving the paradoxes through psychological explanation of the phenomena.

The internal rupture within the Background manifest in various ways in schizophrenia; in what follows I shall divide up the issues in three ways, although it is important to recognise that these are simply different sides of the same conceptual coin. One perspective charts what could be considered the 'ontological' fragmentation of the Background, viewing psychotic symptoms as involving a *dissociation of criteria* for the attribution of propositional attitudes, criteria that normally hang together (see section 2 below). The co-presence of such criteria is presupposed by our everyday interpretative attitude toward the sane subject; the schizophrenic subject, by contrast, embodies characteristic dissociations of criteria that a) makes the straightforward attribution of various psychological states problematic, and b) makes for the creation of distinctive symptom patterns (especially formal thought disorder). (Such an 'ontological' focus perhaps captures best what I have called the 'schizophrenia intuition'.) The second perspective (section 3) is experiential, and traces the impact of a disintegrating Background on schizophrenic experience – both perceptual (auditory hallucinations) and mental (the experiences of introspective alienation). Here the aim is to understand schizophrenic self-ascriptions without viewing them as breakdowns in an introspective mechanism. The third perspective (section 4) views the partial disintegration of the Background from an *epistemic* point of view. Symptoms such as delusions are viewed as local expressions of a *failure in our epistemic 'contact with reality'*, a disruption of the lived certainties of life and of our praxical engagement with the world which grounds our epistemic schemes. The focus here is not on delusions *per se* but on delusionality / psychotic lack of insight / 'madness'; the latter notion is argued to be conceptually more primitive, delusions being understood as delusional beliefs, other phenomena as delusional experience, delusional thought and so on. Delusion(ality) is such a central concept for the understanding of schizophrenia that it would be remiss to not examine existing non-cognitive characterisations of delusion. The chapter ends with a critique of current formulations, which aims to build on their insights yet avoid their attendant difficulties along the way, in order to arrive at an appreciation of what it really means to say that someone has lost touch with reality.

## **2. Ontological Fragmentation in Schizophrenia**

### i. John Hyman's Treatment of Blindsight as a Template for Understanding Schizophrenia

A helpful model for understanding schizophrenic ego-fragmentation can be found in John Hyman's treatment of the condition known as *blindsight*. It is helpful in three ways: firstly the form of the elucidation offered – blindsight as the dissociation of the criteria for vision - can be employed to represent characteristic schizophrenic disorders of thought, affect and action. Secondly the elucidation reveals how and in what way psychological understanding can arise without psychological explanation. Thirdly and relatedly it respects the paradoxical nature of the condition.

'Blindsight' is the paradoxical name given to what appears to be a paradoxical neurological condition, in which subjects sincerely claim that they are blind, or that they can see very little, but act in a way which seems to betray a significantly intact visual system. Weiskrantz's patient D.B., for example, who had his right occipital lobe surgically removed, was found by normal tests (which ask the subject whether and what they can see) to be absolutely blind in the left half of his visual field, but to be able to accurately reach out with his hand towards visual stimuli presented in this area.<sup>208</sup> D.B. himself was astonished by his ability, claiming to be totally unaware of the stimuli and just randomly guessing at their location.

Whilst neurological explanations may go some way towards providing an understanding of the condition, they will hardly be able to remove the paradox and provide the *psychological* understanding needed to make the condition intelligible. It is only natural then to expect that psychological explanations will be offered in order to satisfy the urge for psychological understanding, and Weiskrantz and colleagues were not slow to provide them.

The principal such explanation suggested that the neurological lesions were paralleled by a psychological 'lesion' or 'disconnection' or dissociation. Normally the ability to react to the perceptual stimuli goes hand in hand with a monitoring system which 'creates conscious experience' and provides us with 'a form of privileged access' to private experiences. (Whether the monitoring system is supposed to create or report on these private experiences is not made clear). But in the case of D.B., Weiskrantz

---

<sup>208</sup> Weiskrantz, *Blindsight*.

argues, the monitoring system has become disconnected. The action-directed stream of information processing continues, but without the extra level of internal processing directed at this stream which allows for the subject's commentary.

As will be obvious, this monitoring system is nothing other than Frith's 'self-monitoring mechanism', and builds on the mentalist's idea that our capacity to self-ascribe experience is dependent on the exercise of a faculty of inner sense. This faculty, I argued in chapters 3 & 5, was purely mythical, the product of an alienated conception of subjectivity that was unable to model our authority in self-ascriptions and which went against the grammar of 'sensation', 'thought', 'desire' etc. Although it appears to offer the promise of a psychological explanation, this is only achieved by framing the discussion within this vision of a subjectivity fundamentally alienated from itself. The 'explanation' that is provided then is bogus, only the *form* of an explanation, detailing an impossible malfunction within a mythical framework.

None of this critique however serves to diminish the air of paradox surrounding the phenomenon, or provides any alternative way of understanding the condition. John Hyman, however, after providing a detailed critique of Weiskrantz's theory in line with the sketch above, does go on to develop such an understanding, an understanding which provides psychological understanding without offering a psychological explanation.<sup>209</sup> He first notes the following remark of Wittgenstein's on (as it turns out) *understanding*:

The use of the word 'understand' is based on the fact that in the vast majority of cases when we have applied certain tests, we are able to predict that a man will use the word in question in certain ways. If this were not the case, there would be no point in our using the word 'understand' at all.<sup>210</sup>

That is to say, the criteria for understanding a word are various, and include such phenomena as being able to paraphrase a sentence in which the word occurs, being able to define or otherwise explain the word's meaning, being able to use the word correctly, being able to judge whether others use the term correctly, and so on. In the normal run of things human behaviour is such that these criteria will coincide,

---

<sup>209</sup> Hyman, *Visual Experience and Blindsight*.

<sup>210</sup> Wittgenstein, *Lectures on the Foundations of Mathematics*, p. 23.

and our use of the term depends on this coincidence. Nevertheless, this coincidence is just a contingent feature of human life, and things could be different, in which case the preconditions for the stable employment of the concept of 'understanding' would be absent.

The same, argues Hyman, goes for vision as it does for understanding, for (p. 192):

a correct description of what is before one's eyes, the avowal that one can see it, an appropriate affective response to a visible threat, pointing, indicating whether it is light or dark, simply walking with confidence, are all criteria of visual perception.

In the normal run of things these indices go hand in hand, and this 'background constancy of our interactions with the visible world is part of the framework within which perceptual verbs are used'. But again this constant conjunction is just a contingent feature of human life, which could fall apart. And it is Hyman's suggestion that in the case of blindsight the criteria *do* fall apart. Some of the criteria of vision are applicable, but others are not, and it is this which generates the paradoxical description of 'blindsight'. We are irresistibly inclined to ask the question 'Can D.B. see or not?', but this is the one question that we cannot ask given the failure in preconditions for the applicability of the concept. Hyman's account of blindsight makes it clear *why* this question cannot be asked, and thereby reduces our puzzlement at the phenomenon. It doesn't give a psychological *explanation* of blindsight, but it does, through its explication of the paradox, give us an *understanding* of what it means to suffer from the condition.

## ii. Applying the Template to Schizophrenia

To see how this model can be applied to spell out the 'schizophrenia intuition' first consider once again Frith's theory of the kind of schizophrenic poverty of action that may be manifest in a verbal fluency task (CN pp. 46-8; the task is: how many names of animals can you come up with in 3 minutes?). On Frith's scheme, the subject has the *goal* of responding to the request, but this goal fails to get translated into action. This is not because – as in the case of dementia – the subject's *vocabulary* has been reduced (they could, for example, provide the names of pictured animals). What then can be the

explanation? Frith suggests that 'The subject has to perform a self-directed search through their inner lexicon to find words that are members of the designated category. Schizophrenics with negative signs produce abnormally few items. ... Schizophrenics with negative symptoms produce few items because they find it difficult to perform a self-directed search.'

Unfortunately this is not an explanation of anything, but rather a redescription in a mechanistic vocabulary. What we know already is that the schizophrenic, as judged by certain criteria, does not have a diminished vocabulary, and that, again judged by certain criteria, they have every intention of complying with the experimenter's request. Furthermore, they have no motor impairment that would prevent response. Yet, paradoxically, they still do not respond. To explain away the paradox Frith postulates a failure in a self-directed search of an inner lexicon; but to describe the subject's vocabulary (their *practical* knowledge of words) as (p. 48) an 'inner lexicon' (a representational piece of knowledge) through which the subject has difficulty 'searching' is to invoke a purely mythical operation over a purely mythical substrate. All that has been achieved is a recapitulation of the mentalist's age-old category mistake which views the *retention of a capacity* (memory) as the *storage of an item*<sup>211</sup>. (Hence the substitution of 'lexicon' for 'vocabulary'.)

What, I suggest, is required is to psychologically *acquiesce* in the situation - not to produce a psychological explanation but rather just to acknowledge the phenomena as they occur. In some ways the subject has the intention to comply with the request, in some ways they do not - which is not to say that in some ways they intend *not* to comply with the request, but rather to admit that *in some ways* the ascription of the intention is not apt. The normal criteria for the attribution of intentions include for example both avowal and action, and normally these criteria run together - or if they do not then psychological explanation (*qua* excuse) or physiological explanation (motor impairment) will be forthcoming. In schizophrenia however, and especially when considering poverty of willed action, the criteria come apart, in much the same way as they do in blindsight (see section 2.i above). If the criteria in question were always dissociated, then the language game of intention would simply lose its point. Thankfully for our social negotiations they do not, but unfortunately in schizophrenia they do and, I suggest, it is just this dissociation that provokes the characteristic bafflement at the symptoms, which

---

<sup>211</sup> C.f. Roger Squires, *Memory Unchained*.

justifies the 'schizophrenia intuition', and which gives rise to the characteristic schizophrenic *ununderstandability* identified by Jaspers.

The situation with such simple cases as poverty of action (*qua* abulia) is however hardly the most compelling point to spell out the schizophrenia intuition with the descriptive 'dissociation of criteria' thesis. Furthermore – and if the thesis is right then this is to be expected – poverty of action, however *common* a negative sign it is, is hardly a *paradigmatic* symptom. Similarly with flattening of affect: Frith's theory that true 'inner' emotions remain intact and that the schizophrenic dissociation is between the intact emotion and a failed expression (CN, p. 52: schizophrenic patients 'have difficulty in using their faces to express emotion' and are 'also impaired in the use of the voice to express emotion') only gets off the ground by relying on an implausible alienated conception of emotion. On this story emotions do not naturally put themselves on our faces, but rely on us to put them there. Reverse the conceptual distortion however and such expressions once again become  *criterial* for the affects they express; but now, correlatively, the scope for the explanation envisaged is dramatically reduced.

The schizophrenic situation as regards poverty of affect is not however exhausted by reference to a total absence of emotion. There is also the frequent situation that patients displaying no signs of emotion will report, if pressed, a great deal of emotion<sup>212</sup>. It is here that the disjunction of criteria – between linguistic avowal and more general modes of expression – becomes apparent. In some ways the schizophrenic subject can be said to be emotionally flat, in other ways they can be said to be emotionally charged. The focus on the fragmentation of the Background, the falling apart of the normal preconditions for coherent affect-state attribution, allows one to avoid having to explain this situation psychologically, to avoid having to answer the question as to whether the subject is *really* affectively blunted or charged.

The situation with regards emotional incongruity is even more striking: a patient reacts with emotional expressions that are markedly out of place given the tone of the conversation or the nature of the context. Most disconcerting of all perhaps are the facies often found in catatonic states, where a deadpan emotionless face can be accompanied by lively and expressive eye movements. Here the criteria for the attribution of emotional states utterly dissociate, and no coherent predication can be made either way. In such extreme situations it will undoubtedly still be tempting to employ affective concepts – to

---

<sup>212</sup> Sass, *Schizophrenia, Self-Experience and the So-Called "Negative Symptoms"*, p. 155.

describe the eye movements as expressive of emotions, for example. And there is of course no rule against doing so, so long as the context of the ascriptions is not overlooked. If the apparently catatonic subject – whatever their facies – reacts verbally with unexpected emotional sensitivity and avows intelligible emotions, then there may be considered enough of a context to describe the eye movements as expressive of emotion. On the other hand, there may not. Two philosophical considerations are paramount: Firstly, psychological states are what they are only within a certain (Background) context – they pick out certain patterns in the weave of our lives<sup>213</sup>; secondly, the requisite context in schizophrenia may not only be deteriorated, but even include components that would normally contraindicate the ascription in question. The bafflement that this deterioration of constitutive context produces in us needs to be acknowledged, and not diminished by a falsification of the data in one direction (the absence of emotion or intention) or another (the presence of such cognitive/affective states).

The affective symptom that Bleuler considered most pathognomic of schizophrenia, not yet discussed (and not mentioned at all by Frith), is affective ambivalence. A natural description of this condition ascribes to the subject with schizophrenia both a positive and a negative attitude towards a person, idea or action. Of course we all have mixed feelings for people and towards projects, but this usually produces a unitary ‘hedged’ feeling, and does not present as a straightforward duality. On occasion too we may sincerely avow a positive feeling but our behaviour may betray a negative feeling that is possibly dynamically unconscious (i.e. repressed). Such situations are however necessarily the exception rather than the rule: it is because our utterances and our actions generally hang together in a coherent fashion that we earn the right to be treated as culpable capable emotional agents.

The person with schizophrenia however manifests both positive and negative attitudes equally clearly, or at least, this description of their expression comes naturally to mind. Equally, however, the label does not itself fully describe what is so striking and disconcerting about the phenomenon. For what we are presented with is not a straightforward ambivalence, but what seems to be an actual co-presence of feelings. The contrast can be brought out more clearly when considering intellectual ambivalence, another important schizophrenic symptom. We may all at times be in two minds about something, but

---

<sup>213</sup> Wittgenstein, *Zettel* §§ 532-3: ‘The concept of pain is characterised by its particular function in our life ... we only call ‘pain’

what has *this* position, *these* connections.’



this does not mean that we are actually prepared to both accept and reject, give assent to and dissent from, this topic or concern. We should say: 'Well, on the one hand X, but on the other hand not X, so I don't know what to think', and not: 'X and not X'. Similarly with feelings: these may at times be mixed, but this indicates an affective lack of resolve, and not what is found in schizophrenia, namely, the apparent co-presence both positive and negative attitudes.

The disorientating ununderstandability of this symptom can best be understood, I would suggest, not by trying to reconcile the disparate signs, not by providing the subject with schizophrenia with one 'true' emotion and viewing the appearance of ambivalence as just that – an appearance generated by a defect in expression. But neither will it do to simply ascribe both positive and negative feelings to the subject with no more thought about it than that, for it is in the very nature of feelings to possess a certain degree of consistency. Just as the ascription of ideas is governed by Davidson's constitutive principle of rationality (ch. 1, section 3.i), the ascription of feelings is governed by what could be called a 'constitutive principle of affectivity'. If this seems to make the description of the schizophrenic emotional situation an impossible affair, then all well and good, for unless this paradoxicality is captured then the psychotic quality of the symptom will simply have been missed. On this account it is precisely *because* the ambivalence in question is of *such a deep kind that it actually fragments the phenomena which it characterises* that it can be counted as a psychotic or schizophrenic symptom.

As with blindsight, then, key schizophrenic symptoms can best be understood in terms of a dissociation of the criteria that normally go together and holistically pave the way for the ascription of psychological states. In describing schizophrenic symptoms it is natural to use terms derived from ordinary folk-psychology – derived that is from a situation in which the conditions required by the constitutive principles of rationality, affectivity, conation etc. are met. But it is important to realise that the normal conditions for the use of such terms are not met in the schizophrenic situation, and this because of the depth of the dissociations in question. In fact, talk of 'dissociations' is at once misleading, if it is taken (as it normally is) to refer to dissociations of faculties or cognitive capacities and not of criteria for the ascription of faculties; for it is such faculties etc. that constitutively fragment and which, because of the holistic constraints involved in their predication, cannot in any case be individuated

piecemeal.<sup>214</sup> Nevertheless, however misleading the description, the fact should be noted that it comes naturally to describe schizophrenic symptoms in such terms (to talk of ambivalence etc.), whatever the strict improprieties of doing so. And the reason is clear – given that *many* of the criteria are present that would normally sanction the ascription of some particular emotion or thought or intention, even though others are absent or others even contraindicate the ascription, the ascription is made, and perhaps negated at the same time. The natural terminology of schizophrenia is a function, one might say, of the attempt to apply the normal intentional stance to the psychotic subject and to balk at the difficulties encountered along the way.

The above treatment of affect and intention could doubtless be extended further, to cover ambivalence of will and characteristic aspects of formal thought disorder. The above presentation however is only intended to provide the form of a psychological theory, and not to supply the details. But even in its present state it helps to render intelligible what can often be the obscure theories of schizophrenia provided by the phenomenological traditions. Consider for example Conrad's gestalt analysis of schizophrenia which charts the various phases of those acute schizophrenic shifts which can at times be considered as progressive degrees of deterioration.<sup>215</sup> (The character of the psychotic experience of reality in the *stimmung* will be described further in section 3 below.) First of all in the *stimmung* (i.e. in the 'trema') the coherence of the schizophrenic's lived world is viewed as becoming loosened, whilst later (i.e. in the 'apophany') the coherence is so loose that the gestalt or 'field structure' is lost. Positive formal thought disorders are explained in terms of a progression of lack of gestalt structure, starting with thought broadcast, continuing with thoughts heard allowed, and finally into hallucinatory voices. Similarly in perceptual understanding: delusional perception is characterised in terms of a loosening of perceptual coherence, and in a more severe phase the coherence of perception breaks up altogether (in what Conrad calls the apocalyptic phase). Such a disruption results in catatonic

---

<sup>214</sup> Recognising this allows for a rational reconstruction of Jaspers' critique of the Burgholzli clinicians (Bleuler, Jung etc.) who were prone to model schizophrenia along the lines of hysteria – in other words, as involving the splitting off of individually intact portions of the psyche (the dissociation of 'complexes'). As with the cognitive theories criticised in the previous chapters, such an approach is liable to underestimate the depth of the psychotic disturbance, modelling schizophrenia in terms only applicable to neurosis or normality.

<sup>215</sup> *Die beginnende Schizophrenie. Versuch einer Gestaltanalyse des Wahns*, cited in Fish's *Schizophrenia*, pp. 157-161.

forms of schizophrenia where only fragments of sense experience remain. (A final stage ('apocalypse') leads the patient into an acute and deadly catatonia.)

The point of this description is not to endorse Conrad's analysis, but rather to suggest that the gestalt structures in question can be given a positive characterisation in terms of the diverse criteria for psychological states. Such criteria naturally cluster and co-occur, thereby providing for the possibility of applying our ordinary folk-psychological vocabulary. In schizophrenic states the criteria dissipate, and the gestalts become progressively more deteriorated until they are lost altogether.

At a certain fairly advanced stage of deterioration it is not surprising that symptoms are found which are hard to classify one way or another. It is not just that it is difficult given the disorder of the psychotic mind to *evaluate* whether a patient is thinking or feeling this or that; we do not have to do with a merely *epistemic* problem. Rather there is in the fragmenting mind a lack of the very structures which sustain the application of not only ordinary folk-psychological categories, but in the end even psychopathological vocabulary. Hence Forel's patient L.S. who commented that "It is often impossible to make an exact distinction between delusions, illusions and hallucinations."<sup>216</sup> Similarly, it is frequently impossible to draw any clear distinction between auditory hallucinations and thought insertion in schizophrenia: the hallucinated voices are often described as being un-localised in space, and as being akin to thoughts. Finally, typical psychotic thoughts themselves often indicate a disjunction of criteria for their ascription. At times the schizophrenic subject appears to express a perfectly coherent thought, and we feel at ease with them, feel that they are understandable. But then a minute or two later one is forced to revoke the ascription when the same content suddenly appears in an utterly delusional context, or when it is straightaway contradicted by the patient, or when a response to a question about the thought which asks for clarification manifests a considerable degree of thought disorder. Just as in blindsight we are unsure and unsettled about whether to say that the subject has such and such a perception, so in schizophrenia it can be equally unclear – in a constitutively undecidable way – whether we should say that a subject has such and such a thought.

---

<sup>216</sup> Cited by Bleuler, *DP*, pp. 381-2.

### 3. Schizophrenic Self-Estrangement

#### i. Wittgenstein's 'Secondary Sense' as a Template for Understanding Thought Insertion

The cognitive theories examined in chapter 5 aimed to account for the bizarre experiences of hearing voices and of thought alienation (thought insertion, removal, broadcast, blocking) in terms of a faulty mechanism of self-consciousness. An *a priori* argument was given against such theories to the effect that no such mechanism exists, and that postulating one, far from explaining our ordinary capacity to self-ascribe, actually distorts the grammar of self-ascriptions in a way that undermines the authority of the person *qua* subject. (Another argument would be that such a procedure makes it hard to understand the intrinsically delusional quality of thought insertion.) Nevertheless, the schizophrenic symptom of thought-insertion is a real one, and it may seem that in dismantling the cognitivist theory of 'self-monitoring' or 'introspection' we are left without any way of understand the symptom.

Of course no one (except the occultist) suggests that thought insertion is a genuine phenomenon; in this sense at least we are not forced to take the schizophrenic's words as seriously as they themselves do. It is however common in the psychological literature, and tacitly motivated by an honest enough philosophical principle, to remark that the *experience*, if not the theory behind it, must be taken seriously. And this is surely right, for the subject with schizophrenia is not in any sense *pretending*. (The honest philosophical principle is the one documented in chapter 3, of the impossibility of error concerning our own experiences.)

The tendency then is to suggest that it is at least true that it *seems* to the schizophrenic that thoughts are being put into or taken out of their heads. A strong temptation for the psychologist is now to ask what it is that gives rise to this appearance. In other words, the approach taken is to inquire into *what would have to be the case for it to seem to us as if thoughts were being put into our heads that were not our own*. Once the question has been raised in this way the inclination to suggest that there must be some self-monitoring mechanism or a faculty of self-consciousness within which a disturbance has arisen is practically irresistible. It is just this however which distorts both the philosophical grammar (first-person authority) and the psychiatric phenomenology (delusional nature of the symptom).

What is needed is an account of thought-insertion which shows how the patient's self-ascriptions can be respected and taken seriously without being taken literally. At the same time it is important to note that the patient is not speaking metaphorically; at least, they are not aware of so doing, and it isn't clear that any more literal paraphrase of the self-ascription could be forthcoming.

A clue as to how to understand thought insertion self-ascriptions can be found in part II of Wittgenstein's *Philosophical Investigations*, which contains remarks on aesthetics and on what might be called the 'musicality of language' which rarely receive as much attention as the remarks on mind and normativity found in part I of the book. The relevant remarks which concern what Wittgenstein calls *secondary sense* have been helpfully discussed by Oswald Hanfling<sup>217</sup>; the following is based on his analysis. A useful feature of this analysis is that, as with Hyman's elaboration of Wittgenstein's remarks on understanding, the way is paved for a psychological understanding which does not involve psychological explanation – and which will not, therefore, run the risk (taken by the cognitivist) of producing phenomenological or logical distortions of the data.

Consider the following (*PI* p. 216):

Given two ideas 'fat' and 'lean', would you be rather inclined to say that Wednesday was fat and Tuesday was lean, or the other way round? (I incline to choose the former.) Now have "fat" and "lean" some different meaning here from their usual one? – They have a different use. – So ought I really to have used different words? Certainly not that. – I want to use *these* words (with their familiar meanings) *here* .... Whatever the explanation, – the inclination is there.

Asked "What do you really mean here by 'fat' and 'lean'?" – I could only explain the meanings in the usual way. I could *not* point to the examples of Tuesday and Wednesday.

Here one might speak of a 'primary' and 'secondary' sense of a word. It is only if the word has the primary sense for you that you use it in the secondary one.

---

<sup>217</sup> Hanfling, '*I Heard a Plaintive Melody*'. Secondary sense is an important component of the Background – of that which underpins our understanding of one another and our capacity to self-ascribe (c.f. Part 1 above).

Now Wittgenstein is not concerned to give an *explanation* of this phenomenon ('I say nothing about the causes of this phenomenon'); such an explanation might perhaps be given by neurology or physiology or, as Wittgenstein uncommittedly suggests, childhood associations. He is concerned rather to avoid false explanations and ultimately *just to note that the phenomenon occurs*. For example, restricting the concept of 'metaphor' to cases in which what is said metaphorically can be re-phrased literally<sup>218</sup>, he suggests that

The secondary sense is not a 'metaphorical' sense. If I say "For me the vowel *e* is yellow" I do not mean: 'yellow' in a metaphorical sense, - for I could not express what I want to say in any other way than by means of the idea of 'yellow'.

To generalise this: the use of the word in a secondary sense to make some point is not simply a prosaic way of saying something that could be spelled out otherwise. In this context it is particularly striking that we are compelled in the way that we can be to employ the secondary sense, and that we tend to agree in our uses. Furthermore, whilst there may be *causal* explanations for the choices of words made, these are not rationalising explanations, and even though there are no such rationalising explanations to be had, it is nevertheless true that we are consistently drawn to use such words in such ways.

The examples given so far are perhaps rather peripheral to everyday uses of language, but Wittgenstein (and Hanfling) go on to examine several other more pertinent cases.<sup>219</sup> 'High' is a term used to describe relative altitude, but it is also used to describe musical notes. 'Sharp' and 'flat' also have secondary roles within music. 'Deep' and 'shallow' describe holes and sounds but also feelings and responses. 'Sweet' is used to describe tastes, but also smiles and certain children. These last couple of examples concern the human soul, and it is Wittgenstein's contention that much of the vocabulary we use to describe thought and feeling operates via secondary sense. This is to say both that such terms and phrases should not be taken literally (as the cognitivist takes them, with their mentalist literalisation of the concept of the 'inner' and mechanist conception of 'mental processes'), and that they cannot be

---

<sup>218</sup> Contrast the use of 'metaphor' by Lakoff & Johnson (cf. fn. 7).

<sup>219</sup> The best collection and treatment of such cases can be found in Lakoff & Johnson's *Metaphors We Live By*.

spelled out literally - nor put otherwise - nor rationalised. (If we want to say what they say then we have to use just these words – or perhaps other terms which also operate via secondary sense.) Further examples given by Hanfling include the names given to, and the adjectives used to describe, sensations: ‘pins and needles’, ‘butterflies in the stomach’, ‘stabbing’, ‘grinding’ and ‘burning’ pains.

## ii. Applying the Template

An example which is particularly pertinent for this dissertation on alienation is given by Wittgenstein in his *Remarks on the Philosophy of Psychology*<sup>220</sup> - the example concerns a ‘feeling of unreality’ which he sometimes experienced. The mood in question demanded to be described as one in which ‘everything seems somehow not *real*’, but this is not to say that it is ‘as if one *saw* things unclear or blurred; everything looks quite as usual’. There are no behavioural criteria for the ascription of this mood: the same mood is ascribed to others *simply on the basis of their inclination to use the same words*. The groundlessness of the description is stressed further by Wittgenstein here:

But why do I choose precisely the word ‘unreality’ to express [this mood]? Surely not because of its sound. (A word of very like sound but different meaning would not do.) I choose it because of its meaning. But surely I did not learn to use the word to mean: *a feeling*. No; but I learned to use it with a particular meaning and now I use it spontaneously like *this*.

This feeling of unreality is of course a characteristic aspect of the schizophrenic *stimmung* and therefore merits attention in its own right. What is important is not that *no* explanation can be given as to why Wittgenstein was inclined to use the word ‘unreal’ in the present case. Some such explanation may indeed be forthcoming, and I shall consider one in section ii. below. The important point is that no such explanation will constitute a *rationalisation*; the explanation of why Wittgenstein describes his experience as he did might make reference to certain causal factors, but it will not be rationalised in a any literal manner by referring to features of his experience. If he *could* justify why he used language in the way he did then it would no longer be the case that the criterion for the identity of the experience is that it

---

<sup>220</sup> Volume 1, p. 125. Cited in Hanfling *op cit* p. 129.

is aptly described by the expression in question. We would no longer be dealing with a case of secondary sense.

A similar logical feature, I believe, characterises much of the self-ascriptions of a subject with schizophrenia. They have (or *at least* had) mastery of the ascription conditions for thoughts, the language-games of containment and of the introduction of objects into spaces. The usual secondary-sense metaphor of the mind as an inner realm embodied in our surface-grammar (which the cognitivist misguidedly literalises) has been mastered by them. This mastery, too, is not simply a matter of learning a finite number of folk-psychological verb and noun forms; the metaphor organises and gives form to vast swathes of grammar and makes for the construction of new sentences readily comprehended by others who share the language.<sup>221</sup> And then, given this mastery, they suffer a psychotic episode which involves experiences which can only be described as of someone else putting thoughts into my head which are not my own, or have experiences of their own body moving which can only be described as ... (made movements), or ... (made affects).

The strengths of this kind of account are numerous. On the one hand it is not an objection to argue that metaphor is something that is consciously engaged in and that the person suffering schizophrenia is not intending to talk metaphorically. The metaphors in question are not of this sort: when I say that I have something *in* mind, or something *on* my mind, or that I am feeling *down*, or that things are looking *up* for me, I am not intending to talk metaphorically. I may not even know how to spell out such thoughts differently: the uses of 'in' our 'down' here are fundamental and are not stand-ins for a more literal presentation. On the other hand what is offered is an account which not only shows why psychological explanation is not *necessary* for psychological understanding, but which also shows why such a form of explanation is in fact *undesirable*. I am not thinking merely of the unfortunate *epistemological* consequences already outlined which develop from supposing that a rationalisation for such self-ascriptions (with the consequent invocation of criteria of correctness for such ascriptions and therefore of the possibility of error which would negate the constitutively inalienable character of our subjectivity) is

---

<sup>221</sup> The classic development of this theme – of the structural rôle of metaphor operating at the very heart of our language – especially our folk-psychological discourse – is provided by Lakoff and Johnson in *Metaphors We Live By*, see also Mark Johnson's *The Body in the Mind*.



desirable. What is perhaps even more important is the *psychopathological* advantage of being able to accommodate thought insertion as both an experience *and as a primary delusion* – i.e. as a phenomenon which, again constitutively, *cannot be rationalised without losing its delusional status*.

Another advantage can be found in considering the character of history-taking for first-rank symptoms. What is acknowledged to be important is to not put words into the subject's mouth; to rather have them hit on spontaneous expressions for what is going on in their mind. But most important are the non-rationalising explanatory vistas opened up for coming to terms with what is particularly psychotic and disturbing about the first rank symptoms. This comes from recognising that secondary sense and metaphor are not peripheral features of our lives and discourse, but in fact *foundationally structure* much of our experience, thought, understanding and language use. There is no space in this dissertation to defend this claim<sup>222</sup>, but the content of it is clear. Our mastery of secondary sense is implicated right to the roots of our sensibilities, and this because it forms an essential part of the Background – it constitutes a precondition for our mastery of folk-psychological idioms and our ability to understand the psychological lives of others.

If, then, it is the Background that ruptures in schizophrenia, the sheer strangeness of psychosis is not so difficult to understand. Psychosis strikes right to the heart of our sensibilities – provoking a radical alienation from the grounds of everyday understanding. And this is because it strikes at those sensibilities which are not the *consequence* of our folk-psychological capacities but are rather its *preconditions*. Any normal rationalising psychological account would simply import the Background and tacitly use it as a backdrop against which any particular psychological phenomenon is to be understood. If what psychosis disrupts is the Background itself it is not difficult to see why any explanation which subsumes schizophrenia within normal psychology is doomed to failure.

If secondary sense and metaphor do have the foundational rôle in understanding and language use which I have suggested other schizophrenic symptoms become intelligible. Neologisms are one particularly striking example. If pre-intentional sensibilities (rather than internalised rules) govern our use of many terms then an alteration in such sensibilities through a fragmentation and/or realignment of the Background would naturally result in new uses of terms or the coining of new terms (expressive of

---

<sup>222</sup> For defence and development see the works in the above footnote.

alien sensibilities).

According to some theorists – and Wittgenstein’s treatment of secondary sense might include him in this category – our secondary sense is a brute irreducible fact about us. Others however (especially the linguistic theorists Lakoff and Johnson, and, in a quite different idiom, the Kleinian psychoanalysts) stress the rôle of *bodily experience* in the development of Background sensibilities. Johnson for example suggests that ‘propositional content is possible only by virtue of a complex web of non-propositional schematic structures that emerge from our bodily experience’, or in more detail that:

The centrality of human embodiment directly influences what and how things can be meaningful for us, the ways in which these meanings can be developed and articulated, the ways we are able to comprehend and reason about our experience, and the actions we take. Our reality is shaped by the patterns of our bodily movement, the contours of our spatial and temporal orientation, and the forms of our interaction with objects.<sup>223</sup>

An alteration then in that bodily experience which forms the bedrock of our understanding of meaning, of our action, and of our reasoning may well lead to the kind of schizophrenic symptoms that emerge at levels above that of bodily experience – in delusional symptoms more generally.

Such symptoms will be the focus of the following section. For now I wish merely to note that lived experiences of embodiment are frequently disturbed in schizophrenia. Kinaesthetic hallucinations, often linked with delusions of alien control, are not infrequent. Relevant hallucinations include not only sensations felt in the limbs and other proprioceptive abnormalities, but also interoceptive abnormalities in the guts, extraordinary genital sensations etc., all often with striking delusional elaborations. Disorders of body image or body schema are frequent (especially in prodromal stages) and a bodily basis to disturbances in the boundaries of the self is often manifest.

Stressing the bodily basis of those experiences which are presupposed by the integrity of the self may be not only psychopathologically but also psychotherapeutically relevant. After all our muscular and sensory bodily orientation is not a simple physiological given, but is capable of manipulation: the effects not only of exercise on our sense of bodily well-being, but of dance, breathing, yoga, Alexander technique, Feldenkrais method, and in particular the bodywork therapies (bioenergetics, rolfing etc.) on

---

<sup>223</sup> Quotes from p. 5 and p. xix of *The Body in the Mind*.

not simply the flexibility of the body but on the emotions, on sexual response and fundamentally on our sense of identity are well known. Reichian theorists have long suggested that in neurotic anxiety disorders repressed affects are manifest in functionally identical chronic muscular tensions, and that directed massage can therefore be of use in unblocking the expressed affects<sup>224</sup>. Any masseur will provoke tears on a regular basis, tears which are not always explicable to the client. Alexander Lowen has even suggested that the splits within the schizophrenic personality are directly observable in those aspects of character manifest in posture, claiming to be able to diagnose the condition on this kind of observational basis.<sup>225</sup>

Whatever the scientific credentials of bioenergetic therapy, or the philosophical credentials of Reich's notion of the 'functional identity' of muscular tension and repressed affect, the thesis of the bodily grounding of the Background and the corollary thesis of schizophrenia as a partial fragmentation of the Background are suggestive for the possibility of bodywork intervention in schizophrenia. In fact the thesis of the foundational rôle of lived bodily experience in the Background provides a more satisfactory account of the relation of mind to body than Reich's crude identity thesis: the logico-categorical differences are respected by a theory which allows mind an emergent character – rather than one which paradoxically locates an affect with a particular intentional content in a particular muscular portion. If bodywork therapy can realign bodily experience, promoting the integrity and balance of the lived body, and if the thesis of the bodily basis of Background is on the right lines, then effects on *emergent* features of mindedness – on agency, intentionality, rationality – would naturally be expected.

#### **4. Delusion: Epistemological Alienation in Schizophrenia**

##### **i. Jaspers and *das Wahnproblem***

The ontological focus of section 2 above – on the splits internal to the psyche – requires supplementation by an epistemological orientation that considers the splits that occurs between the

---

<sup>224</sup> Wilhelm Reich, *The Function of the Orgasm*; Nick Totton, *The Water in the Glass: Body and Mind in Psychoanalysis*.

<sup>225</sup> Lowen, *The Language of the Body*.

psychotic subject and the world. Without this the sense in which a psychotic subject lacks 'contact with reality' cannot be understood. The experiential aspect of this has been considered in section 3, but traditionally the principle locus for psychotic lack of or disruption in contact with reality has been the doxastic state of *delusion*.

Those cognitive theories of delusions investigated in the last chapter were argued to be faulty because the kind of explanation they gave was not geared up properly to the nature of the explanandum. They attempted to understand delusion by rendering its emergence intelligible in terms of the experience of the subject, or in terms of general deficits in reasoning. But the first of these only works for secondary delusions; it cannot explain primary delusions because it is forced to ignore their constitutive irrationality. And the second is not only psychologically dubious (general reasoning deficits are either absent or not very pronounced in schizophrenia, and delusions are particular not general), but conceptually suspect (primary delusions (such as found in delusional perception, and autochthonous delusions in general) are *by definition* not the product of reasoning). In short, cognitive theories attempt to understand delusions by making what is necessarily *ununderstandable* understandable. The question then arises of how we are to understand what it means to be deluded – a difficulty known in the literature as *das Wahnproblem*. Answering this question will involve providing a descriptive account of what Jaspers called the *internal characteristics of true or primary* delusions. It will also involve explaining - what the cognitive theories do not - what it is about delusions that manifests the schizophrenic's fundamental *lack of touch with reality*.

Jaspers' own answer to *das Wahnproblem* was sketchy and in some ways unclear,<sup>226</sup> although compared with other psychopathologists the efforts he made were outstanding. Any account of delusion must at least come to terms with his theories, which will now be outlined. According to Jaspers (*GP* p. 106) 'the criteria for delusion proper lie in the *primary experience of delusion* and in the *change of the personality*.' In what follows I shall first consider this 'primary experience', and in later sections detail various attempts to understand psychotic changes in what Jaspers terms the 'personality'.

## ii. The Intrusive Knowledge of Meaning

---

<sup>226</sup> *GP* p. 95: 'a clear presentation is hardly possible with so alien a happening.'

The first set of criteria (the 'primary experience') relates to the *origin* of delusion, and the second to the *adherence* to the delusion by the psychotic subject. In talking of the *primary experience* of delusion Jaspers was in some ways simply manifesting his phenomenologist's prejudice toward experiential readings of mental life<sup>227</sup>, but was also just theorising negatively, rejecting the views that delusions resulted from a generally defective faculty of reason or as *interpretations* of abnormal experience. The perhaps unfortunate experiential focus is further diminished when the distance from *perceptual* experience is stressed: a delusional experience is (*GP* p. 99) 'an immediate, intrusive knowledge of meaning'.

To understand this intrusive knowledge of meaning it is helpful to sketch the circumstances in which (in schizophrenia) it characteristically occurs. A typically delusional experience is *the delusional atmosphere*, an experience of unbearable *uncanniness*. Perception itself remains unaltered, yet the total environment seems subtly yet pervasively changed in an indescribable way, and this characteristically leaves the incipient schizophrenic feeling distrustful and extremely uncomfortable. (This part of the *stimmung* is known as the *trema* – a kind of 'stage-fright' as it were.) There may be a sense that the world is in some way unreal, alien, of a cold glassy smoothness; or that it is false, pretend, with objects like stage accessories and people like puppets. Or perhaps objects may be experienced not in terms of their function, but as mere existents which have lost their meaning. Relatedly, objects may no longer be seen in terms of larger meaningful wholes, but appear as discrete and disconnected. Sometimes following on from (but sometimes contemporaneous with) such experiences is an experience of *apophany*, which is to say, an inchoate sense of new meaning – the belief that something – *something* – unusual is going on.<sup>228</sup>

To gain a sense of the *trema* consider walking round a ghost town (something I did in south west Turkey). Without thinking about it one naturally expects to see people coming and going in out and of

---

<sup>227</sup> As also in *GP* p. 97: 'phenomenologically [delusion] is an experience'. (The grammar of delusion is surely more akin to that of belief than experience: they persist when they are not being entertained by the subject; expressions of delusions ('the world is about to be destroyed') have the form of judgements and seem to express beliefs; furthermore, the diagnosis of delusion is not made pending an investigation of an origin within experience.)

<sup>228</sup> *GP* pp. 98-104; also Louis Sass, *Madness and Modernism*, ch. 2. esp. pp. 46-51. The common descriptions of the *Stimmung* are couched in *metaphor* and rely upon *secondary sense* – see section 1.ii and 3 above.

houses, commerce to be occurring, lights to be on. Instead there is an eerie silence; nothing in particular is striking by its presence, but the sheer absence of life and meaning leaves a considerable impression. Now imagine that the everyday world of the normal town produces the same effect: this is the *trema*, an experience not characterised by perceptual illusion or hallucination, but by a changed sense of the everyday which has suddenly lost its familiarity.

When considering our everyday appreciation of the environment, Jaspers distinguishes (*GP* pp. 94-5) between '*immediate certainty of reality* and *reality-judgement*.' Reality-judgements are 'the result of a thoughtful digestion of direct experiences. These are tested out against each other; only that which stands the test and is confirmed in this way is accepted as real.' By contrast, our immediate certainty of reality is a primary experience of existence (an 'awareness of Being'); it is what is lost, for example, in depersonalisation and derealisation. I unreflectively have a sense of my bodily position and expect the ground to meet my feet at a certain time when walking. A subject with a phantom limb, however, may attempt to step on it and fall: their immediate certainty of reality is distorted.

In relation to this scheme, Jaspers considers that whilst secondary delusions may be distorted *reality-judgements*, primary delusions always imply (p. 95) 'a transformation in our total awareness of reality', that is, involve an alteration in our *immediate certainty of reality*. According to Jaspers, this immediate certainty of reality is distorted in the delusional atmosphere in such a way that (p. 99) 'the environment offers a world of new meanings.' Normal perception comes with a perception of meaning: a house is to live in, 'people in the streets are following their own pursuits. If I see a knife, I see a tool for cutting.' And the experiences of primary delusion 'are analogous to this seeing of meaning, but the awareness of meaning undergoes a radical transformation. There is an immediate, intrusive knowledge of meaning and it is this which is itself the delusional experience.' So delusional significance can be taken on by objects, as in delusional perception: there are Spanish and Turkish soldiers in the streets; the whole town is going to be demolished (to give two of Jaspers' examples). Delusions of reference (p. 100) are also explained in this way: 'Gestures, ambiguous words provide 'tacit intimations'.'

The account Jaspers gives of what he means by a 'seeing of meaning' is not terribly clear, but some rational reconstruction might be possible. To give an example: when someone I meet offers their hand toward me, I automatically take this as a greeting and as a sign that I should grasp it in a handshake. This

is not a 'reality-judgement' on my part; it is not something that I decide on the basis of consideration; I immediately sense the intention behind the outstretched hand. A delusional person, by contrast, might immediately sense this gesture as a threat, or as a 'tacit intimation', as a sign or code. *This understanding is not reached on the basis of thought; it rather reflects an immediate way of understanding the gesture.* Similarly, gestures or communications with particular meanings may be seen where gesture or communication, let alone a particular intention, is not a consideration. (Meanings seen in number plates, design seen behind a random spacing of plants, intention seen behind coughs and sneezes.)

Correlative with this altered experience of meaning is a change in what Jaspers terms the 'personality', which is to say, the (p. 428) 'characteristic totality of meaningful connections' which necessarily form a whole, or a 'total context' for the attribution of intentional and preintentional states. This change is required by an understand of delusion, for delusions are not simply fleeting impressions but are rather firmly maintained beliefs with (p. 196) a 'distinctive incorrigibility'. Nevertheless, it has to be admitted that the character of the change of the personality – the change in the subject's way of looking at the world (as manifest in their values, expressions, ways of loving, self-conception etc.) – is nearly (p. 105) 'impossible to describe, let alone formulate into a concept.'<sup>229</sup> What Jaspers does say is that the personality change is a changed way of looking at the world, in some ways pervasive – a changed sense of reality, but in some ways local, leaving the general capacity for judgement intact. In what follows I shall critically consider several different modern approaches to understanding this change in the personality.

### iii. Change in the Personality 1: Spitzer's Formulation

Whilst Manfred Spitzer is himself one of the chief architects of DSM he is not unaware of the conceptual problems that are contained within some of its formulations. Many of these problems as they

---

<sup>229</sup> In retrospect, and if the positive account of schizophrenia set out below is correct, Jaspers' difficulties are not surprising – for the requisite developments in the theory of knowledge required to understand delusion did not occur until Wittgenstein's *On Certainty* (1969), reliable interpretations of which were not available until Marie McGinn's *Sense and Certainty* (1989) and Avrum Stroll's *Wittgenstein and Moore on Certainty* (1994).

relate to the delusion are helpfully reviewed by him in his paper *On Defining Delusions*. Spitzer's own solution which focusses on the distinctive incorrigibility of delusion noted by Jaspers involves defining delusions as (p. 391) 'statements [sic] about external reality which are uttered like statements about [i.e. avowals of] a mental state, i.e. with subjective certainty and incorrigible [sic] by others.' In effect the first two of Jaspers' 'external criteria' are adopted, the third criterion (impossible or false content) is dropped, and the requirement that delusions concern external reality is added.

In short, Spitzer aims to explain the incorrigibility of delusions by means of an analogy with avowals of mental contents: the delusion is *just like* such an avowal, only it concerns *external* reality. Two problems arise from this definition, one of which (concerning the external restriction) Spitzer attempts to meet head on, the other of which (concerning the avowal comparison) is not mentioned by him. To discuss these in reverse order: the chief difficulty with the definition is that it *doesn't* seem to genuinely explain the distinctive incorrigibility of delusions. Avowals of sensations *are* quite different, both logically and phenomenologically, from avowals of (what can be extensionally identified as) delusions. Expressions of delusions are not believed by delusional subjects to be about themselves (unlike avowals); reasons may be given by a delusional subject for their beliefs but no reasons can be given for sensation self-ascriptions, etc. But even if such discrepancies could be avoided, it still seems that analogy, and not explanation, is all that we are given.

The other problem in the definition comes from the restriction of delusions to external reality. On the one hand this restriction is necessary given the avowal comparison, as otherwise all first-person psychological ascriptions would, ridiculously, have to be considered delusional. But on the other hand it makes it difficult to see how the label 'delusion' can be applied to those feelings and thoughts and experiences and moods which have traditionally been viewed as delusions. Spitzer proposes to reclassify such phenomena as disorders of experience.

The trouble with such a proposal is ultimately not, perhaps, that it forces us not to classify the delusional atmosphere (e.g.) as a delusion, a classification which – given the prototypical status of delusions as *beliefs* – was perhaps always a little strained. Rather, it encourages us to overlook the sense in which the key to what is genuinely psychotic about delusion may be best expressed by the adjective and not the noun. That is to say, we could restrict the term delusion to *delusional* beliefs, but the



commonality of beliefs thus described with delusional feelings, delusional thoughts, delusional experiences, delusional atmosphere, delusional mood etc. should not be overlooked. Talk of mere *disorders* of experience does not help to preserve the sense in which all such disorders seem to indicate psychosis - *not* just in the sense in which they commonly co-occur in conditions that we choose to refer to as 'psychosis' 'madness' 'mania' 'schizophrenia' etc. – but rather in the sense in which *all* such symptoms indicate *a lack of contact with reality and a particularly fundamental disorganisation of the psyche*.

The claim being made is that key symptoms of schizophrenia (for example) are not best understood as psychotic because they tend to occur in such and such constellations. This is the impression given by the recent literature on psychiatric nosology and in the DSM itself, which has largely abandoned the psychotic/non-psychotic distinction. Such literature, nosologically driven by aetiology and not phenomenology, has tended to stress the importance of focussing on particular symptoms rather than on syndromes. Whilst this may make sense when the focus of attention is on research into genetics and on cognitive-behavioural therapeutic intervention, it is (I would argue) less than helpful when considering phenomenological and psychological questions about what it means to suffer from schizophrenia. The tacit assumption of recent psychiatric and cognitive psychological literature seems to be that notions like 'out of touch with reality', 'psychotic', 'delusional' etc. are more or less vague metaphors which can be dispensed with when a more scientifically motivated attention is paid to particular 'cognitive deficits'.

Two mutually supportive considerations help in deciding the issue, one coming from an epistemological and the other from a psychiatric perspective. First, it is hardly surprising that cognitive approaches to psychosis are unable to find the requisite commonality amongst the fundamental symptoms of schizophrenia which would justify more than a summative approach to nosology – hardly surprising given the epistemological orientation of cognitivism. To recapitulate some themes from the first half of this dissertation: Cognitivism supposes that our contact with the world is a matter of accurate *representation* (in perception and thought) and correct *belief*. Knowledge for example is commonly conceived of as a special kind of *belief* – topped up with add-on ingredients (such as justification and truth and some elusive something-else). Our active involvement with the world is seen as merely *consequent* upon independently conceivable inner representational states. Given this restriction in scope

of the notion of our contact with the world to the merely representational it is not surprising that little in common can be found between classically psychotic symptoms (delusional atmosphere, delusional thought, delusional belief, delusional perception). And given this epistemological orientation it is not surprising that nosological categories ('schizophrenia' etc.) begin to look like arbitrary collections of independently conceivable symptoms, and that, presupposing again the validity of such an epistemological outlook, the validity of such diagnoses (or 'constructs' as they might aptly seem to be labelled) is called into question.<sup>230</sup>

Secondly, as Fulford has argued, however much psychiatry shies away from its own fundamental categories such as 'psychosis' that, given its epistemological orientation, it finds embarrassing, the categories in question still play a central rôle not only in psychiatric practice but also, tacitly, in the classification schemes themselves.<sup>231</sup> This indispensability in turn indicates that it may be unwise to write off such (admittedly vague) terms as 'psychotic' or 'out of touch with reality' as not designating relevant and important categories, and perhaps even gives grounds for rejecting an epistemology which fails to locate our fundamental epistemic engagement with reality as that which psychosis disrupts. Admitting this however still leaves us with the responsibility for analysing the psychotic core.

### iii. Change in the Personality 2: Fulford, Campbell and Eilan

To return to the discussion: Jaspers, it will be recalled, suggested that the essence of delusion (or as it might better be put given his treatment of *Wahn*, of delusionality) is to be found both in the primary delusional experience but also in a *change in the 'totality of meaningful connections' which constitute the personality*. A suggestion similar but more detailed than this is provided by Fulford in a paper that draws on John Searle's theory of intentionality. The totality of meaningful connections is called by Fulford (p. 188) a 'background structure of meaning' which 'embeds' the delusional belief in question. It is due to

---

<sup>230</sup> As in Boyle's *Schizophrenia: A Scientific Delusion?*

<sup>231</sup> Fulford, *Moral Theory and Medical Practice* p. 196. Also *Closet Logics...* p. 227. The concept of psychosis may have disappeared from the principle classificatory axes, but it persists as a subcategory of a wide variety of disorders, and in specific categories such as 'psychotic disorder not elsewhere classified'; also in clinical parlance (as in 'puerperal psychosis', 'antipsychotic drug' etc.).

disruptions in the background context to a belief, a background which includes (p. 190) both a 'Network of other Intentional states and a Background of pre-Intentional capacities and stances to the world' – and also (p. 190) the 'connections between beliefs (and other Intentional states) and action' – that the belief in question is to be understood as delusional. The focus that Fulford provides is on the *intentionality* rather than simply the *content* of the delusional belief, a focus which (p. 191) helps explain both i) the fact that delusional beliefs seem somehow 'inherently misdirected' and involve a 'break-down in the relationship between themselves and the world', and ii) also the fact that delusions are easy to identify even though semantically obscure and difficult to describe. (With a delusional patient we find our intuitive ability to employ the intentional stance to be thwarted, a situation which may account for the 'praecox feeling' experienced by a psychiatric interviewer.)

Jaspers' and Fulford's suggestions tell us *where* to look for the solution to the problem of delusion, but they do not in themselves tell us *what* the solution is. As Fulford says (p. 191) 'suggestive as these possibilities may be, the approach, as described here, remains only an approach. It is a sketch for a framework for a pilot project'. What is needed is some way of homing in on the particular disruption that delusions represent within the Background, a description of the particular form or forms that the disruptions in intentionality take, and an explanation of what it is about such disruptions that marks them as delusional or psychotic.

More recently, John Campbell has proposed something of a solution to this problem. According to Campbell the delusional nature of delusions is to be explained in terms of the rôle they play within what we might call the rational economy of the subject<sup>232</sup>. Unlike Spitzer, who's attempt to explain delusions drew an epistemological *analogy* between delusions and folk-psychological self-ascriptions, Campbell proposes an *identity* between delusions and certain fundamental *framework beliefs* that ground our epistemic schemes.

The notion of a framework belief is owed (in content if not name) to Wittgenstein who came to consider the nature of our most fundamental beliefs when elaborating a response to scepticism in his work *On Certainty*. There are various different ways in which the notion of framework beliefs can be explicated, but the following collects some of the principle ideas. Framework beliefs are our most

---

<sup>232</sup> Campbell, *Delusions*; a paper given in 1999 in the department of experimental psychology at Oxford.

*general* beliefs ('the world contains physical objects', 'the sky is above us', 'the sky is sometimes blue'), they are *not questioned* in practice, they speak more of our *understanding of the words* we utter ('here is one hand' said by Moore whilst holding up his hand) than of our capacity to make empirical judgements. They are *not* easily understood as *falsifiable* or *eliminable* or *revisable* or *justified* or *unjustified* because, amongst other things, to doubt their correctness would simultaneously be to doubt the understanding of the believer of what it is that they believe. Framework beliefs concern that of which we are *certain*. They form the *epistemic fulcrums* around which adjust our other beliefs about the world; specific sceptical doubts can only intelligibly be raised against the background of certainty provided by such beliefs.

According to Campbell, then, the delusional of delusions is to be explained by their playing for the subject a rôle in their epistemic economy similar to that which framework beliefs play in the economy of a sane subject. They are experienced by others as unusual and opaque because they cannot be treated in such a way by others. As Naomi Eilan says<sup>233</sup>:

We can, as we say, fall in, to an extent, with a deluded subject's reasoning. In doing this we attempt, precisely, to let a primary belief function as a framework belief. But this is something we cannot actually sustain precisely because we cannot treat it as more than a restricted hypothesis. We cannot put it on a level with all our real framework beliefs, and adjust the latter and our specific reasoning accordingly. Hence the sense of wild unpredictability for us of the way in which, for the schizophrenic, other claims get absorbed in the new framework.

Campbell's proposal seems to succeed with respect to understanding delusions in two respects. First, it helps spell out the sense in which the subject with schizophrenia can be said to sometimes occupy a different 'world' than the non-schizophrenic subject, and explains what it is in the 'personality' or 'totality of meaningful connections' that is altered in schizophrenia. Second, it explains the peculiar intransigence of the psychotic subject, why they are unwilling to give up their delusional beliefs, the way in which seemingly contradictory evidence becomes assimilated into or modified by the delusional system, and so on. This, then, spells out Jaspers' concern, that an alteration in the structure of the

---

<sup>233</sup> Eilan, *On Understanding Schizophrenia*, p. 109.

personality must be posited in order to understand the persistence and immunity to counter-argument of the delusion.<sup>234</sup>

Nevertheless, whilst the proposal makes sense of the *intransigence* of the delusional subject, it does not in itself explain the *bizarreness* of delusions, does not reveal what it is about delusions which makes them so incredibly strange and unbelievable. For it does not seem that this strangeness is captured simply by reference to the intransigence. There is a sense in which we really *cannot* understand delusions. Herein lies the weakness of Eilan's suggestion that we cannot treat a delusion as 'more than a restricted hypothesis', for in most cases, unless we put ourselves into a highly disengaged frame of mind (in which case we forgo an understanding of what it *means to believe the delusional belief*), *we cannot even do this*. Whilst some of our inability to understand delusions may relate to factors outside of the analysis of delusion (Eilan, for example, plausibly suggests that we shy away from attempting to understand certain delusions because the attempt would be too emotionally unbearable<sup>235</sup>), Jaspers' original intuition is that delusions *essentially* involve a particular kind of strangeness manifest in the particular form of ununderstandability they possess. What is required is an explanation of why the different 'world' that the schizophrenic occupies is not simply different but in some way *impossible*; why it is that they do not merely have a *different* way of being in touch with reality, but have in some fundamental way *lost* touch with reality.

#### iv. On the Road to the Psychotic Core: The Praxical Foundations of Delusion

One way in which this can be imagined is by comparing the schizophrenic's predicament with that of the solipsist or sceptic. The aim of much of Wittgenstein's later philosophy was to reveal that such philosophical positions, whatever their apparent intelligibility when cloaked with certain pictorial devices, metaphors, ways of thinking and forms of feeling, pieces of philosophical rhetoric etc., *are actually self-undermining*. They help themselves to concepts and procedures in order to make their

---

<sup>234</sup> GP, p. 411.

<sup>235</sup> Eilan *op cit.* pp. 110-113.

claims, concepts and procedures that should not be allowed given the claims that are made<sup>236</sup>. (The solipsist helps himself to concepts such as 'self' for which nevertheless he is unprepared to supply intelligible contrasts; the sceptic feels content to call into question the background against which any intelligible doubt must be framed.) Now the schizophrenic's predicament is of course far more severe than that of the existentially perturbed philosopher: the philosopher's concerns are second-order, and whatever *intellectual* predicaments they find themselves in, their capacity to think and act and believe and doubt and understand other minds is not affected in practice, when they emerge from their stove room or go play a few rounds of billiards. But with the schizophrenic the praxical grounds of understanding are themselves perturbed. The beliefs they espouse are in some sense impossible. And understanding this helps explain why the schizophrenic's estrangement from the world is at the same time an internal fragmentation or self-undermining of the psyche.

One of the reasons why the altered framework beliefs of the schizophrenic are in some sense impossible<sup>237</sup>, which explains why we can't believe them, is that to believe them would involve doubting our own framework beliefs. A psychotic subject believes that they have an army of people in their mouth; they believe that time has stopped; that the people around them are robots that have no sensations; that the sun is speaking to them. There is a sense in which we just don't know what it would be to have such beliefs, because to believe them would involve doubting that which we can't doubt: we can see the progression of events, we know that something larger cannot fit inside something smaller, we see the emotions and hear the thoughts that others have on their faces and behind their words, we know that speaking requires a mouth etc. To really attempt to believe the schizophrenic would not just, as Eilan suggests, require us to adopt an existentially or emotionally unbearable stance, but require us to stop believing that which rationally structures our world, that which makes for the possibility of our having a

---

<sup>236</sup> To employ Derrida's pretty phrase, their conditions of possibility are the same as their conditions of impossibility.

<sup>237</sup> In *On Certainty* Wittgenstein sometimes contrasts mistakes with delusion in a way which brings out the inadequacy of the cognitivist's conception of delusion as simply false belief. (§71 If my friend were to imagine one day that he had been living for a long time past in such and such a place, etc. etc., I should not call this a mistake, but rather a mental disturbance, perhaps a transient one. Cf also §257: If someone said to me that he doubted that he had a body, I should take him to be a half-wit. But I shouldn't know what it would mean to try to convince him that he had one. And if I had said something, and that had removed his doubt, I should not know how or why.')

'world', that which provides for our cognitive purchase on the world in the first place. If we were to renounce the world in the attempt to empathise with the schizophrenic, and to succeed in this renunciation, empathy would still be impossible as our minds would fall apart in the process.<sup>238</sup>

The focus so far has been on those certainties of our lived worlds that are characteristically expressed in *belief*. But whilst this enables capture of something of the sense of delusions, that is, of delusional beliefs, it hardly allows for an understanding of other delusional phenomena (delusional atmosphere, perception, thinking), nor brings the account of delusion any closer to that of hallucination and such characteristically psychotic phenomena as thought insertion. Furthermore it fails to mine the depths of Wittgenstein's critique of scepticism. For as *On Certainty* progresses, the certainties which Wittgenstein believes to lie at the foundation of our world are revealed to be not *propositions* believed with certainty, but rather certainties expressed in *action*<sup>239</sup>. This praxical focus is manifest in such paragraphs as §204 'it is our *acting* which lies at the bottom of the language-game'; §359 'I want to conceive of [certainty] as something that lies beyond being justified or unjustified; as it were, as *something animal*. [my italics]'; §342 'certain things are *in deed* not doubted'.

Such certainties make reference not so much to my knowledge but rather to my confident action, my fluent self-expression and self-understanding, my capacity to read the feelings of others. I am certain that when I put my foot down the floor will resist me, that I can make myself more or less understood to my friends, that I can judge distances approximately by sight. These are not certainties that are ever normally formulated by those who are certain in the ways described: my confidence about the floor resisting my foot is manifest in my stride and balance, and not in any kind of judgement I make.

These praxical certainties are, as Wittgenstein puts it, beyond being justified or unjustified. They provide the ground for our conceptual schemes, and in the most general sense define what it is to be a rational animal. In so far as delusional beliefs contradict our framework beliefs, we cannot argue against them by providing some kind of *justification* for our cleaving to such framework beliefs. (The sceptic,

---

<sup>238</sup> I am intending here to echo various Kleinian psychoanalytic theories of schizophrenia (psychosis as self-attack on mindedness, a la Bion, for example) without taking a stance on the motivational dynamics (defensive splitting etc.) that the analysts claim to detect.

<sup>239</sup> C.f. Avrum Stroll, *Wittgenstein and Moore on Certainty*, p. 7 and pp. 146ff.

too, errs in demanding a justification for our fundamental beliefs, errs in supposing that it is unreasonable of us to have such beliefs without justification. To be sure, justification and reasons cannot be provided, but then they are not required – and could not be required. Our spade is turned, bedrock has been reached, when the praxical preconditions of the language-game are what is being brought into focus. The sceptic wants these preconditions to feature as moves within the language-game.) And in so far as delusional beliefs function as abnormal framework propositions, they occupy an impossible rôle. In a sense, there is no *reason* why we shouldn't have such framework beliefs. All that can be said is that to have such beliefs is *not to reason*, not to be a rational animal. Being rational, thinking, being the kind of creature with respect to whom the intentional stance can be deployed, involves, in the final analysis, thinking like *this*, believing like *this*. The rationality involved is not a matter of correct inference; it is, as it were, the rationality of our animal nature.

Consider the *Stimmung* in this light. The world becomes for the schizophrenic a terrifying place. Something uncanny seems afoot. Nothing seems quite right. Objects seem to have lost their familiar meanings and rôles. Some of this can be quite literally understood as a disintegration of the capacity to view events and the gestures of Others against the meaning and intelligibility-conferring Background that normally gives them sense. The holistic context presupposed by language and other social artefacts which gives them their sense is not automatically adopted, and in this light, it is the physical non-semantic properties of objects and words that gain ascendancy. Furthermore, the uncanniness and terrifying nature of the world as it appears in the *Stimmung* can be viewed in terms of the absence of the lived certainties that ground our normal rational engagement with the world, an absence of the epistemic Background.

It is against the background of the terrifying experience of the *Stimmung* that primary delusions can crystallise out (in the apophany). In the light of the above discussion, such delusions can be seen as the finding of new certainties out of the epistemic chaos of the *trema*, new grounding patterns of action, response and thought that generally employs an already known vocabulary but which is sometimes in the process driven to employ neologisms. These certainties of lived action, however, are not simply alternative ways of going on, or rather, whilst at the level of mere 'going on' they are simply alternatives,



they are not alternatives which give rise to the possibility of those structures constitutive of mindedness, rationality and the distinctive human form of life.

#### **v. Subpersonal Dynamics and the Rehabilitation of Cognitive Neuropsychology**

Giving a psychological explanation of such a transformation will always come up against the limits imposed by the category of psychosis itself – as that which represents a breakdown in the conditions of possibility for the provision of psychological explanation in the first place. This is especially so if cognitive explanation is attempted, for it is the cognitive faculties – at least, those pertaining to what the psychoanalysts call the ‘secondary processes’ – which suffer most from the psychotic disintegration. More primitive aspects of mindedness – the primary processes – various emotions, on the other hand, may still be unproblematically ascribable and allow for some kind of non-rationalising explanation (of the general form provided by psychoanalysis). None of this however precludes the provision of neurophysiological explanation, and it is at such a sub-personal explanatory level that some of Frith’s cognitive theories criticised in chapters 4 & 5 can be rehabilitated.

Talk of ‘rehabilitation’ is perhaps rather derogatory; on the one hand, and as already repeatedly stressed, the critique of philosophical cognitivism developed in parts 1 & 2 was not a critique of cognitive psychology itself. On the other hand, cognitive science involves a loose and evolving set of theories with different protocols, paradigms, and different means of adequation of concepts and verification of claims continually providing different senses to the questions which the science raises. A certain ‘natural selection’ operating within Frith’s theory of self-monitoring itself seems to have occurred over the years, the content of the hypotheses now being directed less at the psychological and personal level and more at the neurological and distinctly subpersonal level of inquiry, the level from which the examples of corollary discharge in the visual system operating in the theory as analogies themselves functioned.

Speaking generally it is undeniable that the possibility of coherent action and the possibility of adequately ‘tracking’ the world is underpinned by a large variety of interconnected feedback mechanisms. These neurophysiological mechanisms operate between subject and world, providing

central or peripheral feedback between bodily movement and sensory stimulation, or operate solely within the body. It is important to stress the *sub-personal* level of these feedbacks: the neurophysiological categories do not map neatly onto personal level categories (of perception, action, and intentional content).<sup>240</sup> The emergent (psychological) properties of a self-organising dynamic system (the body, especially the nervous system) *are* truly emergent and cannot be neatly correlated with the structural system properties.<sup>241</sup> An example of such a feedback/feedforward mechanism would be the corollary discharge device mentioned by Frith which makes for the accurate perception of motion and stasis by an organism who's perceptual organs may themselves be moving. Another example would be the neural systems that allow for different reactions to self-produced vocalisations than to other-produced vocalisations. (We are not deafened by our own shouting, but if someone else were to shout at the same distance from our ear as our own mouth is, it would be experienced as deafening.)

A nice example provided by a student of Frith's (Sarah-Jayne Blakemore) is that of the *tickle* response. It is a well-known fact that, whilst we cannot really tickle ourselves, the touch of others may be unbearably tickly. Such a response is explained in outline by Blakemore in terms of a simultaneous feedforward of the hand motor commands to the neural systems implicated in bodily sensation. If a delay between action and sensation is mechanically introduced in our self-tickling, the ticklishness of our own actions is greatly increased, a phenomenon which would be predicted by the feedforward model. The interesting and relevant concern is that patients who heard voices and had passivity experiences did not seem to have this feedforward; self- and other-induced tickles were experienced as having the same intensity.

If such sub-personal self-organising dynamic feedback systems do become disrupted in schizophrenia, the experiences of the *Stimmung* are only to be expected. The praxical certainties that ground our experience and thought will be partially undermined to the extent that the subpersonal systems which make for the possibility of such certainties will be undermined. In so far as the

---

<sup>240</sup> As Susan Hurley argues in *Consciousness in Action* p. 400, 'perceptual distinctions or invariants can depend noninstrumentally on output. Not only are creatures .. who perceive and act essentially situated in environments. Perception and action are also co-constituted.'

<sup>241</sup> C.f. Andy Clark, *Being There: Putting Brain, Body, and World Together Again*.

physiological systems which incorporate or link up various feedback subsystems are forced into new levels of equilibrium, different emergent (i.e. psychological) properties are only to be expected. And because there is no one-to-one correlation between sub-personal properties and emergent personal-level properties, it is easy to see how cognitive psychological theories which seem to presuppose some such modular equivalence will be doomed to failure. Not only this, but the means for explaining the essential, constitutive, ununderstandability of the psychotic symptoms is itself brought into clearer focus. Furthermore, because local defects at the sub-personal level can have unpredictable and non-local effects at the personal level, and because defects in different parts of the subpersonal system may have similar effects on the emergent structures, sense can be made of the fact that disorders with *different* aetiologies and *different* (behaviouristically defined) symptomatologies will still *all* be classified as schizophrenic.

This dissertation has for the main been concerned to trace the inherently doubtful strains in the cognitive psychological theories, to expose the alienated conception of mind they seem to presuppose and the consequent negative implications for the explanatory scope of these theories. But if the positive thesis of schizophrenic symptoms being aptly understood as a function of the disintegration of the Background is on the right lines, then the development of cognitive theories which theorise such Background deficits from a truly *sub*-personal perspective can only be welcomed.

## Bibliography

- American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders III*, 1980.
- American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders IIIR*, 1987.
- American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders IV*, 1994.
- Armstrong, David, *A Materialist Theory of Mind*, Routledge 1968.
- Arrington, Robert & Glock, Hans-Johann (eds.), *Wittgenstein's Philosophical Investigations*, Routledge 1991.
- Austin, J. L., *A Plea for Excuses*, *Proceedings of the Aristotelian Society*, vol. 57., 1956-7, pp. 1-30.
- Bakan, David, *Clinical Psychology and Logic*, *The American Psychologist*, December 1956, pp. 650-660.
- Baker, Gordon P. & Hacker, Peter M. S., *Wittgenstein: Meaning and Understanding*, Blackwell 1992.
- Barron, Dorit & Long, Douglas C., *Avowals and First-Person Privilege*, *Philosophy and Phenomenological Research*, vol. LXII, 2001, pp. 311-335.
- Bentall, Richard, *From Cognitive Studies of Psychosis to Cognitive-Behaviour Therapy for Psychotic Symptoms*.
- Birdwhistell, Ray, *Kinesics and Context: Essays on Body-Motion Communication*, Penguin 1971.
- Blakemore, Sarah-Jayne, *Monitoring the Self in Schizophrenia: The Role of Internal Models*, in Zahavi, D. (ed.), *Exploring the Self*.
- Bleuler, Eugen, *Dementia Praecox: Or the Group of Schizophrenias [DP]*, International Universities Press, 1950 [1911].
- Boyle, Mary, *Schizophrenia: A Scientific Delusion?*, Routledge 1993.
- Brenner, William H., *Natural Law, Motives, and Freedom of the Will*, *Philosophical Investigations*, vol. 24, pp. 246-261.
- Buss, Arnold, *Psychopathology*, Wiley 1966.
- Button, Graham, et al., *Computers, Minds and Conduct*, Polity 1995.
- Campbell, John, *Delusions*, (unpublished paper 1999).
- Candlish, Stewart, *The Real Private Language Argument*, *Philosophy* 1980, vol. 55, pp. 85-94.

- Canfield, John, *Private Language: Philosophical Investigations Section 258 and Environs*, in Arrington & Glock (eds.) *Wittgenstein's Philosophical Investigations*, pp. 120-137.
- Chapman, L. J. & Chapman, J. P. *The Genesis of Delusions*, in Oltmanns, T. F. & Maher, B. A. (eds.) *Delusional Beliefs*, pp. 167-183.
- Chihara, Charles & Fodor, Jerry, *Operationalism and Ordinary Language: A Critique of Wittgenstein*, in Pitcher, G. (ed.), *Wittgenstein: The Philosophical Investigations*, pp. 384-419.
- Child, William, *Causality, Interpretation and the Mind*, OUP 1994.
- Chomsky, Noam, *Rules and Representations*, Blackwell 1980.
- Churchland, Patricia, *Neurophilosophy: Toward a Unified Science of the Mind/Brain*, MIT Press 1986.
- Churchland, Paul, *Folk Psychology and the Explanation of Human Behaviour*, in Churchland, Paul, *Philosophical Perspectives, 3: Philosophy of Mind and Action Theory*, Atascadero 1989.
- Clark, Andy, *Being There: Putting Brain, Body, and World Together Again*, MIT 1997.
- Conrad, K., *Die beginnende Schizophrenie. Versuch einer Gestaltanalyse des Wahns*, Thieme 1958.
- Cooke, John W., *Human Beings*, in Winch, P. (ed.), *Studies in the Philosophy of Wittgenstein*, pp. 117-151.
- Coulter, Jeff, *Approaches to Insanity: A Philosophical and Sociological Study*, Martin Robertson & Co. Ltd. 1973.
- Crow, Tim, *Positive and Negative Schizophrenic Symptoms and the Rôle of Dopamine*, *British Journal of Psychiatry*, vol. 137, 1980, pp. 383-6.
- Crow, Tim, *Nuclear Schizophrenic Symptoms as a Window on the Relationship between Thought and Speech*, *British Journal of Psychiatry*, vol. 173, pp. 303-9.
- Currie, Gregory, *Imagination, Delusion and Hallucinations*, *Mind & Language*, vol. 15, 2000, pp. 168-183.
- Davidson, Donald, *Mental Events*, in *Essays on Actions and Events*, pp. 207-227
- Davidson, Donald, *Essays on Actions and Events*, OUP 1980.
- Dennett, Daniel, *Consciousness Explained*, Penguin 1991.
- Dennett, Daniel, *The Intentional Stance*, MIT Press 1987.
- Dilman, Ilham, *Science and Psychology*, in O'Hear, A. (ed.), *Verstehen and Humane Understanding*, pp. 145-164.

- Dilman, Ilham, *Psychology and Human Behaviour: Is There a Limit to Psychological Explanation?*, *Philosophy*, vol. 75, 2000, pp. 183-201.
- Dreyfus, Hubert L. & Hall, Harrison (eds.), *Heidegger: A Critical Reader*, Blackwell 1992.
- Eilan, Naomi, *On Understanding Schizophrenia*, in Zahavi, D. (ed.) *Exploring the Self*.
- Einstein, Albert, *The Meaning of Relativity*, Methuen 1922.
- Evans, Gareth, *The Varieties of Reference*, OUP 1982.
- Fischer, Eugen, *On the Very Idea of a Theory of Meaning for a Natural Language*, *Synthese*, vol. 111, 1997, pp. 1-16.
- Fish, Frank, *Fish's Schizophrenia*, 2<sup>nd</sup> edition, John Wright & Sons Ltd., 1976.
- Fodor, Jerry, *Psychological Explanation: An Introduction to the Philosophy of Psychology*, Random House 1968.
- Fodor, Jerry, *The Language of Thought*, Thomas Y. Crowell 1975.
- Fodor, Jerry, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, MIT Press 1987.
- Fodor, Jerry, *Could there be a Theory of Perception?*, *Journal of Philosophy*, vol. 63, 1966, pp. 369-380.
- Freud, Sigmund, *The Interpretation of Dreams*, Hogarth Press 1900.
- Frith, Christopher, *The Cognitive Neuropsychology of Schizophrenia [CN]*, Lawrence Erlbaum 1992.
- Fulford, K. W. M., *Minds and Madness: New Directions in the Philosophy of Psychiatry*, *Philosophy*, vol. 37 (supplement), 1994, pp. 5-24
- Fulford, K. W. M., *Moral Theory and Medical Practice*, Cambridge University Press 1989.
- Fulford, K. W. M. *Closet Logics: Hidden Conceptual Elements in the DSM and ICD Classifications of Mental Disorders*, in Sadler, J. Z. et. al. (eds.), *Philosophical Perspectives on Psychiatric Diagnostic Classification*.
- Garety, Phillipa & Hemsley, David, *Probabilistic Judgements in Deluded and Non-Deluded Subjects*,  
 Garety, Philippa, *Reasoning and Delusions*, *British Journal of Psychiatry*, vol. 159 (supplement 14), 1991, pp. 14-18.
- Garety, Philippa & Freeman, Daniel, *Cognitive Approaches to Delusions: A Critical Review of Theories and Evidence*, *British Journal of Clinical Psychology*, vol. 38, 1999, pp. 113-154.
- George, A. (ed.), *Reflections on Chomsky*, Blackwell 1989.

- Glock, Hans-Johann, *A Wittgenstein Dictionary*, Blackwell 1996.
- Glover, Jonathon, *The Philosophy and Psychology of Personal Identity*, Penguin 1988.
- Golberg, Bruce, *Meaning and Mechanism*, in Hyman, John (ed.), *Investigating Psychology*.
- Gold, Ian & Hohwy, Jakob, *Rationality and Schizophrenic Delusion*, *Mind & Language*, vol. 15, 2000, pp. 146-167.
- Graham, George & Stephens, G. Lynn [G&S] (eds.), *Philosophical Psychopathology*, MIT Press 1994.
- Graham, George & Stephens, G. Lynn, *Mind and Mine*, in Graham & Stephens (eds.) *Philosophical Psychology*.
- Graham, George & Stephens, G. Lynn, *When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts*, MIT Press 2000.
- Grice, H. P., *The Causal Theory of Perception*, *Aristotelian Society*, supp. vol. 35, 1961, pp. 121-152.
- Hacker, P. M. S., *Davidson on First-Person Authority*, *Philosophical Quarterly*, vol. 47, pp. 285-304.
- Hacker, P. M. S., *Insight and Illusion*, 2<sup>nd</sup> edition, Thoemmes Press 1997.
- Hacker, P. M. S., *Insight and Illusion*, 1<sup>st</sup> edition, OUP 1972.
- Hacker, P. M. S., *Wittgenstein: Meaning and Mind*, Blackwell 1993.
- Hamlyn, David, *In and Out of the Black Box*, Blackwell 1990.
- Hamlyn, David, *The Psychology of Perception*, Routledge 1957.
- Hanfling, Oswald, *I Heard a Plaintive Melody*, *Philosophy* (supplement 28), pp. 117-133.
- Harth, E. *Windows on the Mind: Reflections on the Physical Basis of Consciousness*, The Harvester Press 1982.
- Haslam, John, *Illustrations of Madness*, Routledge 1988.
- Heidegger, Martin, *Being and Time*, Blackwell 1962.
- Heil, John, *Does Cognitive Psychology Rest on a Mistake?*, *Mind*, vol. XC, 1981, pp. 321-342.
- Hill, P., Murray, R. & Thorley, A. (eds.), *Essentials of Postgraduate Psychiatry*, Academic Press / Grune & Stratton 1979.
- Bob Hinshelwood, *The Elusive Concept of 'Internal Objects'*, *International Journal of Psychoanalysis*, vol. 78, pp. 877-897.

- Hoffman, Ralph, *Verbal Hallucinations and Language Production Processes in Schizophrenia*, Behavioural and Brain Sciences, vol. 9 (3), 1986, pp. 503-17.
- Hornsby, Jennifer, *Physicalist Thinking and Conceptions of Behaviour*, in Hornsby, J., *Simplemindedness*.
- Hornsby, Jennifer, *Bodily Movements, Actions and Epistemology*, in Hornsby, J., *Simplemindedness*.
- Hornsby, Jennifer, *Simple Mindedness: In Defense of Naïve Naturalism in the Philosophy of Mind*, Harvard University Press 1997.
- Hunter, J. F. M., *On How We Talk*.
- Hurley, Susan, *Consciousness in Action*, Harvard University Press 1998.
- Hyman, John, *How Knowledge Works*, Philosophical Quarterly, vol. 49, 1999, pp. 433-451.
- Hyman, John, *The Causal Theory of Perception*, Philosophical Quarterly, vol. 42, 1992, pp. 277-296.
- Hyman, John (ed.), *Investigating Psychology: Sciences of the Mind after Wittgenstein*, Routledge 1991.
- Hyman, John, *The Imitation of Nature*, Blackwell 1989.
- Hyman, John, *Visual Experience and Blindsight*, in Hyman (ed.), *Investigating Psychology*.
- Jacobsen, Rockney, *Wittgenstein on Self-Knowledge and Self-Expression*, The Philosophical Quarterly, vol. 46, pp. 12-30.
- Jaspers, Karl, *General Psychopathology* [GP], Manchester University Press 1963.
- Johnson, Mark, *The Body in the Mind: The Bodily Basis of Meaning, Imagination and Reason*, University of Chicago Press 1987.
- Johnson-Laird, P. N. [J-L], *Mental Models*, Erlbaum 1983.
- Kasanin, J. S. (ed.), *Language and Thought in Schizophrenia*, W. W. Norton 1964.
- Kemp, Roisin (et al) *Reasoning and Delusions*, British Journal of Psychiatry, vol. 170, 1997, pp. 398-405.
- Kenny, Anthony, *The Legacy of Wittgenstein*, Blackwell 1984.
- Kenny, Anthony, *The Homunculus Fallacy*, in Hyman, J. (ed.) *Investigating Psychology*.
- Laing, R. D., *The Divided Self*, Penguin 1990 [originally published Tavistock Publications Ltd. 1960].
- Lakoff, George & Johnson, Mark, *Metaphors We Live By*, University of Chicago Press 1980.
- Leader, Darian, *Freud's Footnotes*, Faber and Faber 2000.
- Lowen, Alexander, *The Language of the Body*, Collier Books 1958.
- MacIntyre, Alasdair, *Against the Self-Images of the Age*, Duckworth 1971.



- MacDonald, G. F. (ed.) *Perception and Identity*.
- MacDonald, C., Smith, B., Wright, C. (eds.), *Knowing Our Own Minds*, OUP 1995.
- Maher, Brendan, *Anomalous Experience and Delusional Thinking: The Logic of Explanations*, in Oltmanns, T. F. & Maher, B. A. (eds.), *Delusional Beliefs*, pp. 15-33.
- Maher, Brendan, *Delusions: Contemporary Etiological Hypotheses*, *Psychiatric Annals*, vol. 22 (5), 1992, pp. 260-8.
- Maher, Brendan, *Delusional Thinking and Perceptual Disorder*, *Journal of Individual Psychology*, vol. 30, 1974, pp. 98-113.
- Malcolm, Norman, *The Myth of Cognitive Processes and Structures*, in Mischel, T. (ed.) *Cognitive Development and Epistemology*.
- Malcolm, Norman, *Dreaming*, Routledge 1959.
- Malcolm, Norman & Armstrong, David, *Consciousness and Causality: A Debate on the Nature of Mind*, Blackwell 1984.
- Marr, David, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W. H. Freeman 1982.
- Maturana, H. R. & Varela, F. J., *The Tree of Knowledge*, Shambhala 1992.
- McDowell, John, *Mind and World*, Harvard University Press 1994.
- McDowell, John, *Criteria, Defeasibility and Knowledge*, *Proceedings of the British Academy*, vol. LXVIII, 1982, pp. 455-479.
- McGinn, Marie, *Sense and Certainty: A Dissolution of Scepticism*, Blackwell 1989.
- Medvydev & Medvydev, *A Question of Madness*.
- Midgley, Mary, *Beast and Man*, Harvester Press 1979.
- Mischel, Theodore (ed.), *Cognitive Development and Epistemology*, Academic Press 1971.
- Morris, Michael, *The Good and the True*, OUP 1992.
- Mullen, P., *The Phenomenology of Disordered Function*, in Hill, P. et. al. (eds.) *Essentials of Postgraduate Psychiatry*.
- Newton, Isaac, *Philosophiae Naturalis Principia Mathematica* [1687]
- Oltmanns, T. F. & Maher, B. A. (eds.), *Delusional Beliefs*, Wiley 1988.

- Parnas, Josef, *The Self and Intentionality in the Pre-Psychotic Stages of Schizophrenia: A Phenomenological Study*, in Zahavi, D. (ed.), *Problems of the Self*.
- Peters, R. S., *The Concept of Motivation*, Routledge 1958.
- Phillips, D. Z., *Epistemic Practices: The Retreat from Reality*, in *Recovering Religious Concepts*.
- Phillips, D. Z., *Recovering Religious Concepts, Closing Epistemic Divides*, Macmillan 2000.
- Pitcher, George (ed.), *Wittgenstein: The Philosophical Investigations*, Macmillan 1968.
- Polanyi, Michael, *Personal Knowledge: Towards a Post-Critical Philosophy*, Routledge 1958.
- Porter, Roy, *Mind-Forg'd Manacles: A History of Madness in England from the Restoration to the Regency*, Athlone Press 1987.
- Prigogine, Ilya, *Order out of Chaos*,
- Proudfoot, Donald, *On Wittgenstein on Cognitive Science*, *Philosophy*, vol. 72, 1997, pp. 189-217.
- Puhl, K. (ed.), *Meaning Scepticism*, de Gruyter 1991.
- Reich, Wilhelm, *The Function of the Orgasm*, Souvenir 1983.
- Rorty, Richard, *Philosophy and the Mirror of Nature*, Blackwell 1980.
- Rundle, Bede, *Mind in Action*, OUP 1997.
- Ryle, Gilbert, *The Concept of Mind*, Hutchinson's University Library 1949.
- Sadler, J. Z., Wiggins, O. P., Schwartz, M. A., *Philosophical Perspectives on Psychiatric Diagnostic Classification*, Johns Hopkins University Press 1994.
- Sass, Louis, *Madness and Modernism: Insanity in the Light of Modern Art, Literature and Thought*, Harper Collins 1992.
- Sass, Louis, *The Paradoxes of Delusion: Wittgenstein, Schreber and the Schizophrenic Mind*, Cornell University Press 1994.
- Sass, Louis, *Schizophrenia, Self-Experience, and the So-Called "Negative Symptoms"*, in Zahavi, D. (ed.), *Exploring the Self*.
- Sass, Louis, *Analyzing and Deconstructing Psychopathology, Theory and Psychology*, vol. 9, 1999, pp. 257-268.
- Schacht, Richard, *Alienation*, Allen & Unwin 1971
- Schroeder, Severein (ed.), *Wittgenstein and Contemporary Philosophy of Mind*, Palgrave 2001.

- Searle, John, *Intentionality: An Essay in the Philosophy of Mind*, CUP 1983.
- Shanker, Stuart, *Wittgenstein's Remarks on the Foundations of AI*, Routledge 1998.
- Shanker, Stuart, *Computer Vision or Mechanist Myopia?* in Shanker, S. (ed.) *Philosophy in Britain Today*, pp. 213-266..
- Shanker, Stuart (ed.), *Philosophy in Britain Today*, Croom Helm 1986.
- Shepard, Roger N. & Metzler, Jacqueline, *The Rotation of Mental Objects*, *Science*, vol. 171, 1971, pp. 701-3.
- Sims, Andrew, *Symptoms in the Mind: An Introduction to Descriptive Psychopathology*, Balliere Tindall 1988.
- Spitzer, Manfred, *On Defining Delusions*, *Comprehensive Psychiatry*, vol. 31, 1990, pp. 377-397.
- Sprague, Elmer, *Persons and their Minds: A Philosophical Investigation*, Westview Press 1999.
- Squires, Roger, *Memory Unchained*, *Philosophical Review*, vol. 78, 1969, pp. 178-196.
- Still, Arthur & Costall, Alan, *Against Cognitivism: Alternative Foundations for Cognitive Psychology*, Harvester Wheatsheaf 1986.
- Strawson, Peter, *Individuals: An Essay in Descriptive Metaphysics*, Methuen & Co. 1959.
- Strawson, Peter, *Perception and its Objects*, in MacDonald, G. F. (ed.) *Perception and Identity*.
- Stroll, Avrum, *Moore and Wittgenstein on Certainty*, OUP, 1994.
- Tanney, Julia, *Playing the Rule-Following Game*, *Philosophy*, vol. 75, 2000, pp. 203-224.
- Tanney, Julia, *Why Reasons may not be Causes*, *Mind & Language*, vol. 10, 1995, pp. 103-126.
- Tanney, Julia, *De-Individualising Norms of Rationality*, *Philosophical Studies*, vol. 79, 1995, pp. 237-258.
- Tausk, Victor, *On the Origin of the 'Influencing Machine' in Schizophrenia*, *Psychoanalytic Quarterly*, vol. 2, 1933, pp. 519-56. [1919]
- Taylor, Charles, *Heidegger, Language and Ecology*, in Dreyfus & Hall (eds.) *Heidegger: A Critical Reader*.
- Taylor, Charles, *Sources of the Self: The Making of the Modern Identity*, Harvard University Press 1989.
- Ter Hark, Michel, *Wittgenstein and Dennett on Patterns*, in Schroeder, S. (ed.), *Wittgenstein and Contemporary Philosophy of Mind*.
- Thornton, Tim, *Wittgenstein on Thought and Language: The Philosophy of Content*, Edinburgh University Press, 1998.

- Totton, Nick, *The Water in the Glass: Body and Mind in Psychoanalysis*, Rebus Press 1998.
- Von Domarus, E. *The Specific Laws of Logic in Schizophrenia*, in Kasanin, J. S. (ed.) *Language and Thought in Schizophrenia*.
- Walker, Chris, *Delusion: What did Jaspers Really Say?*, British Journal of Psychiatry, vol. 159, (supplement 14), 1991, 94-103.
- Weiskrantz, L., *Blindsight*, OUP 1986.
- Wilson, George M, *The Intentionality of Human Action*, North-Holland Pub. Co. 1980.
- Winch, P. (ed.) *Studies in the Philosophy of Wittgenstein*, Routledge 1969.
- Wittgenstein, Ludwig, *Philosophical Investigations [PI]*, 2<sup>nd</sup> edition, Blackwell 1958.
- Wittgenstein, Ludwig, *Remarks on the Foundations of Mathematics*, 3<sup>rd</sup> edition, Blackwell 1978.
- Wittgenstein, Ludwig, *Lectures on the Foundations of Mathematics*, Harvester Press 1976.
- Wittgenstein, Ludwig, *Remarks on the Philosophy of Psychology [RPP]*, vols 1 & 2, Blackwell 1980.
- Wittgenstein, Ludwig, *The Blue and Brown Books*, Blackwell 1958.
- Wittgenstein, Ludwig, *Zettel*, Blackwell 1967.
- Wittgenstein, Ludwig, *On Certainty*, Blackwell 1969.
- Wright, Crispin, *Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics*, in George, A. *Reflections on Chomsky*.
- Wright, Crispin, *Wittgenstein's Later Philosophy of Mind: Sensation, Privacy and Intention*, in Puhl (ed.) *Meaning Scepticism*, pp. 126-147.
- Wright, Crispin, *Self-Knowledge – The Wittgensteinian Legacy*, in Macdonald, C. et. al. (eds.), *Knowing Our Own Minds*.
- Wright, Crispin, *How Can the Theory of Meaning be a Philosophical Project?*, *Mind & Language*, vol. 1, 1986, pp. 31-44.
- Young, & Bentall, Richard, *Hypothesis Testing in Patients with Persecutory Delusions: Comparison with Depressed and Normal Subjects*, British Journal of Clinical Psychology, vol. 34, 1995, pp. 353-69.
- Zahavi, D. (ed.), *Exploring the Self*, John Benjamins Publishing Company 2000.